

A question of integrity

Iran's institutions must investigate allegations of scientific plagiarism as a matter of urgency.

resh evidence that senior officials in the Iranian government have co-authored scientific papers that show signs of plagiarism came to light this week (see page 704). This follows similar revelations in October (see *Nature* **461**, 578–579; 2009).

The first wave of alleged plagiarism cases was widely discussed both inside and outside Iran, and provoked dismay among the country's researchers and reformist bloggers. The cases were also reported by Iran's mainstream media, which deserve credit for airing the story despite the present regime's record of shutting down newspapers, arresting journalists and otherwise intimidating free inquiry.

The regime's research institutions, however, have done little to investigate the allegations. This is perhaps not surprising, given the extreme political sensitivity of the accusations. One of the disputed papers was co-authored by transport minister Hamid Behbahani, who supervised President Mahmoud Ahmadinejad's PhD. Other questionable papers were co-authored by science minister Kamran Daneshjou, who oversaw this year's disputed presidential election. An official investigation is needed to clarify the circumstances of these and the new cases of alleged plagiarism. Senior scientists also have a responsibility for what happens in their labs, and for papers on which their names appear.

One can only speculate over what might cause such plagiarism. In some cases, using texts to help counter a poor command of English may provide extenuating circumstances. And in Iran, as in several developing countries, there is a cultural expectation that officials should have strong academic credentials. This could tempt weaker academics to inflate their publication record, perhaps by plagiarism. It can also lead to other types of academic fraud: in November 2008, the late Ali Kordan was impeached by the Iranian parliament and removed as interior minister after he admitted that his honorary law

doctorate from the University of Oxford was fake.

Another factor could be the politicization of Iran's research system. After the 1979 Islamic revolution, universities were purged of perceived Western influences and staff. Many of the best scientists left the country. The Iranian research enterprise began to recover in the late 1990s under reformist president Mohammad Khatami — academic appointments were more often based on merit, and the country's research output improved in both quantity and quality. But the research environment has deteriorated since Ahmadinejad took power in 2005, as his regime has exerted political influence over promotions within universities. And in the aftermath of this summer's protests over the election, Iran's universities have become a hotbed of opposition — a prime focus for the government's crackdown. The regime now says that it intends to reinforce Islamic values across academic staff and courses, which many scientists take as code for further repression.

Leading researchers inside Iran are keeping their heads down. But many are quietly pressing for the authorities to investigate the plagiarism allegations, which, they note, would be consistent with wider demands by academics for the current regime to be more accountable and respectful of the republic's values and civil rights. They are also pushing for merit-based promotion practices, and are having some success in persuading Iran's academic institutions to emphasize ethics in the practice of research and publishing.

Iran's researchers, both inside and outside the country, are to be applauded for their defence of excellence and scientific integrity in such difficult political conditions. The actions of a few must not be allowed to soil the reputation of the majority of Iran's scientists. Rather, the international scientific community must redouble its efforts to support and collaborate with its Iranian colleagues.

A slippery slope

Animal-research policies should be guided by moral consensus, not by arbitrary decisions.

n 2006, building on its strong veterinary-research programmes, Oklahoma State University (OSU) in Stillwater opened a biosafety-level-3 laboratory to work on high-risk zoonotic diseases and other pathogens. But, this October, OSU president Burns Hargis abruptly cancelled an anthrax-vaccine project that would have used the facilities because the baboon subjects involved would have been killed at the end of the study (see page 706).

Although Hargis has the authority to make such decisions, this action seems arbitrary and ad hoc. The project was approved by its funding agency, the US National Institutes of Health, and by the university's animal-use committee. Moreover, there was nothing untoward in the project's plan to kill the animals; after they have been

infected with high-risk pathogens, they cannot be uninfected.

Speculation was rife on campus that Hargis acted under pressure from Madeleine Pickens, the wife of oil magnate and OSU benefactor T. Boone Pickens. Earlier this year, she complained about how animals were used in surgical training at the university veterinary centre. The OSU, which was to receive a US\$5-million donation from her, has since changed its policy on training procedures.

Spokespeople for Madeleine Pickens and Hargis have denied that she had any role in the anthrax decision. Hargis insists that, despite complaints from faculty members about not being consulted, he did speak to Stephen McKeever, the OSU vice-president for research, and checked the report of the animal-use committee before making his ruling.

Imposing such a decision on the fly sets a bad precedent. Animal-research policies need to be guided by a moral compass — a consensus on what people find acceptable and unacceptable. Precisely because this issue is so morally and emotionally fraught, decisions by administrators should involve careful consultation with researchers as well as with all other members of university communities.

RESEARCH HIGHLIGHTS

When 'wit' is not 'wet'

Proc. R. Soc. B doi:10.1098/rspb.2009.1788 (2009)

Humans can distinguish between similarsounding words regardless of the speaker. Debate has centred on whether this ability is unique to humans or exists in other animals and contributed to the evolution of human speech.

To tackle the question, Verena Ohms of the Institute of Biology Leiden in the Netherlands and her colleagues recorded 21 men and women saying the Dutch words 'wit' and 'wet', then played them back for eight zebra finches (pictured). Enticed by birdseed, the finches learned to discriminate between the two words. The birds could still tell the two apart when the words were spoken by people they hadn't heard before, including people of the opposite sex.



J. BILDARCHIV/ PHOTOLIBRARN

CLIMATE SCIENCE

Carbon sink limits

Geophys. Res. Lett. doi:10.1029/2009GL041009 (2009)

As ecosystems 'fix' nitrogen, converting it to ammonia, they may also be affecting their own ability to soak up carbon.

Ying-Ping Wang of CSIRO in Aspendale, Australia, and Benjamin Houlton of the University of California, Davis, have modelled how the terrestrial biosphere responds to interactions between factors such as nutrient flux, nitrogen fixation and light availability. The team found that ecosystems absorbed less carbon after they ran out of soil nitrogen to fix.

Calculations of future warming and carbon uptake need to take this factor into consideration, the authors say, suggesting that warming could proceed at a faster pace than expected in nitrogen-limited ecosystems.

DRUG DELIVERY

Into the tumour

Cancer Cell 16, 510-520 (2009)

Targeting diagnostic agents or drugs to tumours could make cancer diagnosis and treatment more effective. Researchers have devised ways to get molecules to home in on tumours, but getting them deep into the bulk of the tumour has proved difficult.

Erkki Ruoslahti of the Burnham Institute for Medical Research in Santa Barbara, California, and his co-workers have identified a peptide that they bound to a known cancer drug, as well as to imaging agents, and then injected these into tumour-bearing mice. Magnetic resonance imaging and other methods revealed an 8–11-fold higher level of the peptide-bound drug in tumours than of the drug alone.

The researchers show that the peptide targets and penetrates tumours by first binding to a set of tumour-specific receptors, called αv integrins. It is then cleaved, exposing a motif that binds to another receptor, neuropilin-1, to gain entry to the cell.

IMMUNOLOGY

Worms begone!

J. Exp Med. doi:10.1084/jem.20091268 (2009)
The cells lining the gut combat worm infections by secreting an immune modulator that blocks worms from feeding on host tissues.

Fred Finkelman of the University of Cincinnati in Ohio and his colleagues injected mice with one of two nematode parasites:

Heligmosomoides polygyrus (pictured below) or Nippostrongylus brasiliensis. They

© 2009 Macmillan Publishers Limited. All rights reserv

monitored worm counts, levels of cytokines (immune modulators) and changes in the intestinal lining cells.

The authors found that the cytokines IL-4 and IL-13, known to be produced in response to worm infection, stimulate gut lining cells to secrete another cytokine called RELM- β . This interferes with worm feeding, promoting expulsion of the nematodes from the gut.

PHYSICS

Electron turnstiles

New J. Phys. 11, 113057 (2009)

The ampere is the standard unit of electrical current, but its definition is arbitrary. Physicists would like to redefine the ampere in terms of the motion of a small number of electrons, so that the standard is based on the unchanging charge carried by the fundamental particles.

Ville Maisi at the Centre for Metrology and Accreditation in Espoo, Finland, and his colleagues in Finland and Japan bring such a definition closer. The team uses transistors that act as electron 'turnstiles'—allowing just one of the particles through at a time. Putting ten such gates in parallel generates a measurable current that could potentially be used to define a more precise ampere, the authors say.

NEUROSCIENCE

Hub neurons synch brain

Science 326, 1419-1424 (2009)
In the rodent hippocampus, groups of brain cells link up to each other via highly connected hub neurons. This

may help to synchronize brain activity during development.

Rosa Cossart at the French National Institute of Health and Medical Research (INSERM) in Marseilles and her colleagues imaged neuronal activity in brain slices from young rats and mice. By digitally reconstructing the connections between the neurons, the researchers identified a subset of inhibitory neurons that were highly connected to either nearby or more distant cells.

When stimulated, 8 out of 20 highly connected brain cells and 1 in 25 poorly connected ones seemed to be hub neurons, inducing sustained firing or rhythmic, synchronized activity across the cell network.

PLANETARY SCIENCE

Titan's tub

Planet. Space Sci. 57, 1872-1888 (2009) Beneath the icy crust of Saturn's moon Titan may lurk a watery ocean, say Christian Béghin at the University of Orleans in France and his colleagues.

The researchers examined data from the Huygens probe, which landed on Titan's surface (pictured right) in 2005. On its way down, it measured a surprising electric field signal that, on Earth, is caused by lightning. But no lightning has been recorded on Titan, and the signal is instead interpreted as the result of Saturn's magnetic field interacting with

Titan's atmosphere and interior. The team says that the signal is most consistent with a crust just tens of kilometres thick covering an ammonia-rich ocean.

MICROBIOLOGY

Malaria adapts to host

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0907590106 (2009)

The malaria parasite tends to cause more severe disease in young children. This could be because the parasite expresses genes associated with the severe form of malaria in people with limited immune defences, according to Peter Bull of the Kenya Medical Research Institute in Kilifi and his colleagues.

The researchers sequenced portions of key virulence genes — the *var* genes — from malaria parasites obtained from 217 Kenyan children, 112 of whom had severe disease. Certain *var* genes were expressed at higher levels in younger children and in those with more severe forms of the disease.

The researchers also found that higher *var* gene expression correlated with low levels of antibodies for parasite-infected cells. The team suggests that the parasite can take advantage of low host immunity to cause more severe malaria.

CHEMISTRY

Fuel cells' future

Science **326,** 1384-1387 (2009)

A low-cost nickel-based material could replace platinum as the catalyst that drives the electrolysis of water, the key reaction that powers hydrogen fuel cells.

Vincent Artero at Joseph Fourier University in Grenoble, Serge Palacin at the French Atomic Energy Commission in Gif-

sur-Yvette and their colleagues took a nickel-based catalyst that mimics hydrogenase enzymes and attached it to the surface of multiwalled carbon nanotubes to increase the catalyst's surface area. The researchers then tested the material using a proton-exchange membrane to produce hydrogen from a water-based sulphuric acid solution.

Their system was about 100 times less efficient than a commercially available platinumbased one. But previous enzymeinspired catalysts worked only in organic solvents, whereas this one is stable under real fuel cell conditions, the authors say.



Bird feeder effects

Curr. Biol. doi:10.1016/j.cub.2009.10.061 (2009) Thanks to bird lovers, the central European songbirds called blackcaps (Sylvia atricapilla) have split into two groups that may develop into separate species. So say Martin Schaefer at the University of Freiburg in Germany and his colleagues.

The researchers tracked the groups: one overwinters in the Mediterranean, where the birds feed on fruit; the other spends winter in the United Kingdom, where its members eat from bird feeders. The two groups return to nesting grounds in Germany and Austria at different times, and so tend to pick mates from among the birds that overwinter in the same place. The team found that, in less than 30 generations, the UK-wintering birds have developed distinct physical traits: rounder wings for their shorter migration and narrower beaks, which are less well adapted for fruit-eating.

JOURNAL CLUB
Katherine H. Freeman

Pennsylvania State University
A biogeochemist ponders muddy

A biogeochemist ponders muddy molecules and past climates.

I am amazed by how humble fossil lipids in muddy sediments can yield insight into Earth's history. The structures and relative abundances of these marine biomarkers, which originate from cellular membranes, provide records of physiological and ecological responses to changing ocean chemistry and temperature. They help to quantify ancient climates, and may, for example, offer a peek at the future by providing clues to ocean temperatures when the poles were free of ice.

Shifting abundance ratios of membrane lipids from marine Archaea — a proxy called TEX_{86} — faithfully indicate modern seasurface temperatures. Yet ancient temperatures signalled by TEX_{86} can be significantly higher than those indicated by other proxies, making TEX_{86} hard to interpret.

Julius Lipp and Kai-Uwe Hinrichs at the Center for Marine Environmental Sciences in Bremen, Germany, show that the constituent compounds in TEX₈₆ may be a mixture derived from ancient microbes and those living in muddy sediments today (J. S. Lipp and K.-U. Hinrichs Geochim. Cosmochim. Acta 73, 6816–6833; 2009).

The authors identified the muddwellers' lipids from their polar functional groups; ancient lipids lack these groups because they are quickly lost after burial. The core hydrocarbons waving the polar flags probably account for the proxy's overestimation of temperature. By identifying contributions from organisms living in sediments, the researchers provide a powerful means to discern which environments preserve the primary TEX₈₆ signature and thus under which conditions we can reliably use this important proxy.

Climate scholars should take note and take heart, because this work will ultimately strengthen our interpretations of these muddy molecules to help us better understand Earth's past and future climate.

Discuss this paper at http://blogs.nature.com/nature/journalclub

nature Vol 462|10 December 2009

NEWS BRIEFING

POLICY

Stem-cell lines: On 2 December. the US National Institutes of Health (NIH) approved 13 human embryonic stem-cell lines for use by US governmentfunded researchers — the first lines to be given the green light under a liberalized policy announced by President Barack Obama in March. On 4 December, a standing advisory committee to NIH director Francis Collins recommended he approve an additional 27 lines, provided he limit their use to research projects "consistent with the wording of the consent form" used to obtain the lines. This stipulates that the lines be used to study "the embryonic development of endoderm with a focus on pancreatic formation". A further 96 lines are under review.

Carbon targets: Following a similar announcement by China, India said it would cut its carbon intensity — the amount of carbon dioxide emitted relative to economic output — by 20–25% from 2005 levels by 2020. Environment minister Jairam Ramesh, announcing the commitment on 3 December, said that the target was worked out in concert with other developing countries (see *Nature* 462, 550; 2009).

Legislation denied: Australia's government has for the second time rejected proposed legislation to create a carbontrading scheme. Following a fortnight of political turmoil that saw climate-change sceptic Tony Abbott elected as leader of the opposition, Prime Minister Kevin Rudd declined to call a snap election. The government said that there would be another chance to vote for the scheme when parliament resumes in February. See go.nature.com/ ayfWdf for more.

Unhealthy air: Carbon dioxide and other greenhouse gases are dangerous to human health, the US Environmental Protection



SPACE TOURISM'S NEW ERA

Enthusiastic space tourists got their first public viewing of the commercial passenger vehicle SpaceShipTwo (pictured, centre: mounted under its carrier aeroplane WhiteKnightTwo) on 7 December at Mojave Air and Space Port, California. The rocket ship, developed by aviation designer Burt Rutan and bankrolled by British billionaire Richard Branson, will carry passengers to the edge of outer space — for US\$200,000 a ticket. Virgin Galactic, which Branson hopes will become the world's first commercial spaceline, says it has already signed up more than 300 passengers for flights from a yet-to-be-built spaceport in New Mexico. They are unlikely to get space-borne until 2011 at least — even if all flight testing goes smoothly.

Agency declared on 7 December. The ruling, proposed in April, allows the administration of President Barack Obama to regulate emissions under the federal Clean Air Act without going through Congress. Automobile emissions are likely to be the first to be regulated.

Appeal lost: Italian scientists have lost a final appeal against a government research call that explicitly excludes human embryonic stem cells, even though their use is legal. The researchers objected when the exclusion was added by politicians to a text agreed by a committee of scientific experts (see Nature 460, 19; 2009). They took the health ministry to court in June, arguing that the exclusion infringed a constitutional freedom of scientific research. They lost that case, and on 2 December, they lost their appeal to the supreme administrative court.

SOUND BITES

"It might be in the future we would be forced to lower the prize."

Michael Sohlman, executive director of the Nobel Foundation, thinks that the Nobel prizes might have to be shrunk as a result of the economic downturn.

Source: Reuters

BUSINESS

Carbon capture: On 4 December, the US Department of Energy announced \$979 million for three demonstration projects to capture carbon dioxide from power plants and store it underground. Private investors would add another \$2.2 billion, the agency said. Two projects involve fitting carbon-capture units to existing coal power plants (in New Haven, West Virginia and Mobile, Alabama), and a new power plant will be built near Midland, Texas.

Patent reform: The European Union has inched closer to a system in which a single patent would cover all 27 member states, with just one, centralized appeals court. On 4 December, industry ministers agreed the principles of the plan, which would drastically cut the costs of managing patents across Europe. But they deferred agreement on crucial issues, such as the costly matter of translating

a single patent into different languages. The plan awaits debate in the European parliament.

Biotech boost: The European life-sciences sector was perked up by its first large initial public offering for almost two years, after Belgian biotech firm Movetis raised €85 million (US\$128 million) on its 3 December debut. Backed by venture-capital investment, Movetis has European approval to market a constipation drug, prucalopride (Resolor).

RESEARCH

London medical hub: Details of a new £520-million (US\$850million) biomedical research centre in central London were outlined on 7 December. The UK Centre for Medical Research and Innovation will bring together four research institutes in one building. Paul Nurse, president of the Rockefeller University in New York, who is leading the development of science plans for the complex, said the budget would be "tight"; and operations could begin with fewer than the hoped-for 1,250 researchers if construction costs go over budget. See go.nature.com/ uPkUlY for more.

Climate security breach: Three weeks after the theft of e-mails from the Climatic Research Unit (CRU) at the University of East Anglia in Norwich, UK, unsuccessful hacking was reported at the Canadian Centre for Climate Modelling and Analysis at the University of Victoria in British Columbia.

NUMBER CRUNCH

221

The number of days unmanned underwater glider Scarlet Knight took to cross the Atlantic Ocean, gathering water data. It is the first craft of its kind to make the crossing.

Source: Rutgers, the State University of New Jersey

Andrew Weaver, a climate researcher at the university, claimed that there had been a "sustained hacking attempt" in recent weeks. At CRU, Phil Jones has stepped aside as director, pending the result of a university review of the incident.

Censorship row: Australia's national science agency has been accused of trying to alter a peer-reviewed paper that was critical of carbon-trading schemes, leading a prominent researcher to

quit the agency on 2 December. Clive Spash (pictured below), an ecological economist at the Commonwealth Scientific and Industrial Research Organization in Canberra, said he had "had enough" after his paper was accepted by the journal *New Political Economy* earlier this year, withdrawn by the acting chief of his division, and then approved again on condition that parts were reworded. See go.nature.com/



Tuberculosis funds:

Tuberculosis research has seen funding jump in each of the past few years, but the rate of increase is dropping off. So says a report released on 3 December by the Treatment Action Group, an AIDS research and policy think tank based in New York. Tuberculosis funding increased by 8% last year to US\$510 million, compared with annual increases of 13% in 2007 and 17% in 2006. The balance of funding is also shifting, from government agencies to philanthropic organizations such as the Bill & Melinda Gates Foundation.

Anthropologists in conflict:

A panel of anthropologists has again criticized the Human Terrain System, a controversial US Department of Defense project to embed social

THE WEEK AHEAD

9-11 DECEMBER

New Zealand hosts the Antarctic Treaty Meeting of Experts, which will recommend ways to manage ship-borne tourism to Antarctica.

go.nature.com/cUJGwl

10 DECEMBER

This year's Nobel science laureates receive their awards in Stockholm, Sweden.

go.nature.com/PUfasn

11 DECEMBER

NASA's orbiting infrared telescope, the Wide-Field Infrared Survey Explorer, is scheduled to launch.

http://wise.ssl.berkeley.edu

14-18 DECEMBER

The American Geophysical Union meets in San Francisco, California.

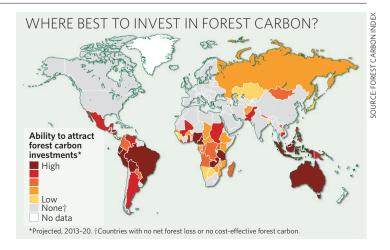
www.agu.org/meetings/fm09

scientists in military units in Iraq and Afghanistan to gain cultural understanding (see Nature 455, 583-588; 2008). In a report presented by the American Anthropological Association at its annual meeting on 3 December, an internal committee said the programme had uneasy ethical tensions — such as being too closely aligned with military intelligence gathering - and could "no longer be considered a legitimate professional exercise of anthropology".

MARKET WATCH

Amazonian nations will be the early winners in any market for forest carbon credits, which could grow to US\$20 billion a year by 2020. The Forest Carbon Index, released by the environmental think tank Resources for the Future and consultancy firm Climate Advisers, both based in Washington DC, charts where governments should invest in preserving forests in developing countries. That is, if the Copenhagen negotiations permit rich nations to offset their emissions this way.

The report says that 85% of the best places for early forest carbon returns (2013–20) are in the greater Amazon, particularly in Brazil and Peru, where there is a high rate of deforestation, cheap land, market capacity and political will to save the forests. Investment will also depend on national policies: Brazil has asked for donations — not an offset market — to help avoid deforestation, because it wants developed countries to reduce emissions at home. But it has recently said it might allow some offsets. See go.nature.com/TtU9Tm for more.



NEWS

Plagiarism scandal grows in Iran

Investigation finds more cases of duplication in publications co-authored by ministers and senior officials.

EXCLUSIVE

Nature has uncovered further instances of apparent plagiarism in papers co-authored by government ministers and senior officials in Iran. The spate of new examples raises questions about whether such incidents are symptomatic of conditions also common in other developing countries — such as difficulties with English or pressure to acquire academic credentials as a prerequisite for promotion — or whether they are also linked specifically to the Iranian regime, where growth of a merit-based university culture has been undermined by political appointments and purges of reformminded scientists (see page 699).

An earlier probe^{1,2} revealed extensive plagiarism in a paper co-authored by transport minster Hamid Behbahani and four papers coauthored by science minister Kamran Daneshjou. The revelations received wide coverage in the Iranian media and blogosphere. Scientists inside and outside the country have called for investigations, as well as for stronger ethical oversight in Iran's research institutions.

Daneshjou, a mechanical engineer at the Iran University of Science & Technology (IUST) in Tehran, was head of the interior-ministry office that oversaw this year's disputed election that kept President Mahmoud Ahmadinejad in power. In October, the Iranian parliament's commission for science and education held an informal inquiry into the four Daneshjou papers. Although it made no official conclusion, it effectively cleared Daneshjou after his co-author, IUST colleague Majid Shahravi, took responsibility for the papers' contents in the Iranian media — although both Shahravi and some members of the commission also maintain that the papers contained originality. Three of the four papers have now been retracted by the journals in question — the fourth was in an Iranian journal.

The paper³ by Behbahani, an IUST researcher who supervised Ahmadinejad's PhD, has not been investigated, although it seems to be almost entirely put together from three earlier articles by different authors². It was retracted by the journal *Transport* in October.

Behbahani has publicly said that the paper did not constitute plagiarism because only parts of the article were identical to earlier work. He challenged the allegations of plagiarism, calling them a "media attack, far from fairness and integrity" and "an illegitimate accusation".



Research papers co-authored by Hamid Behbahani contain text from other works.

Nature has now uncovered yet more instances of apparent plagiarism in papers from Behbahani and some of his co-authors.

One paper⁴ on asphalt-road resistance — by Behbahani's Transport co-authors Hassan Ziari, a deputy minister of roads and transportation whom Daneshjou recently appointed as head of Payame Noor University in Tehran, and Mohammed Khabiri, then a PhD student at the IUST — contains many sections that are identical to a 2005 paper⁵ by scientists in Pakistan.

And two 2008 papers^{6,7} on strengthening asphalt roads, co-authored by Behbahani and Ziari with PhD student Shams Noubakhat, also contain duplicated material. The first⁶ includes multiple passages from three earlier papers^{8–10} and the second⁷ is also largely taken from three other papers¹⁰⁻¹².

One scientist familiar with the field, who asked to remain anonymous, says that he has difficulty making sense of the first paper's results, and that some data in it⁶ are identical to those in one of the earlier papers by different researchers8. "That the two sets of results could be identical is improbable," he says. Behbahani and Ziari did not respond to requests for comment. Muhammad Atif Ramay, managing editor of the Journal of Applied Sciences Research in which both papers were published, says that the journal has withdrawn the articles from its website pending further investigation.

Also in question is a 2008 paper on modelling

pollution in Iran¹³, which is co-authored by one of the 37 members of the Iranian Supreme Council of Cultural Revolution, Mohammad Ali Kaynejad, an environmental engineer at Sahand University of Technology in Tabriz, Iran. The paper almost entirely duplicates a 2001 conference paper¹⁴ on modelling pollution in Hungary.

The Iranian paper acknowledges the original source of the model, although the authors wrote that it was "tested via the simulation of a photochemical oxidant episode that took place in Tabriz, Iran in 2007". But Alison Tomlin, an environmental modeller at the University of Leeds, UK, and a co-author on the Hungary model, says that the Iranian paper contains "no new results" and "is definitely a copy". It includes computer simulations purportedly of Iranian data, but they match the Hungary figures — and the background map outlines Hungary, not Iran.

The first author of the Iranian paper, Esmaeil Fatehifar, an environmental engineer at Sahand University of Technology, places the responsibility on another member of the team. "He said these are measured data about Tabriz Petrochemical Complex," he says. "I thought he was right and accepted it." Fatehifar says he intends to cancel the team member's PhD plans. He adds that Kaynejad had "not seen that paper" even though his name is on it. Kaynejad did not respond to Nature's interview requests.

Questions have also been raised over work co-authored by Ali Reza Ali-Ahmadi, education minister in the previous government of Ahmadinejad. A 2006 paper 15 on supply networks co-authored by him includes many sentences and paragraphs that are identical to those in three earlier papers^{16–18}. Mika Ojala at Tampere University of Technology in Finland, a co-author on one of the earlier studies, says that in his opinion this is not coincidence. Ali-Ahmadi could not be reached for comment. Babak Amiri, an IUST researcher and a coauthor on the paper, says that a draft version of the paper was accidentally submitted before it was checked by himself or Ali-Ahmadi. "I apologize for this big mistake," he says.

Nature has also learned that the US National Academy of Sciences earlier this year removed a chapter from a 2003 book¹⁹ on a US-Iranian workshop. Ironically, the chapter, authored by Hassan Zohoor, secretary of the Academy of Sciences of the Islamic Republic of Iran, was



COPENHAGEN LIVE

Nature's reporters blog from the United Nations climate conference.

go.nature.com/OEsfAa

called 'The impact of moral values on the promotion of science'. It was withdrawn because it substantially duplicated a 1999 paper²⁰ by Douglas Allchin, a historian and philosopher now at the University of Minnesota in Minneapolis.

Zohoor says that he never saw Allchin's paper, and that he only prepared a draft of the paper, leaving others in his office to "develop it and add the literature review". Zohoor says that the explanation of the staff member involved — that the copying happened "quite accidentally and as a mere negligence" — is inadequate, and that he intends to write to Allchin to apologize. "In my entire life I've never copied anyone else's work," says Zohoor.

Declan Butler

- 1. Butler, D. Nature doi:10.1038/news.2009.945 (2009)
- 2. Butler, D. Nature 461, 578-579 (2009).
- Ziari, H., Behbahani, H. & Khabiri, M. M. Transport XXI, 207-212 (2006).
- 4. Ziari, H. & Khabiri, M. M. J. Eng. Appl. Sci. 2, 33-37 (2007).
- Kamal, M. A., Shazib, F. & Yasin, B. J. East. Asia Soc. Transport. Stud. 6, 1329–1343 (2005).
- Behbahani, H., Ziari, H. & Noubakhat, S. J. Appl. Sci. Res. 4, 96-102 (2008).
- Behbahani, H., Ziari, H. & Noubakhat, S. J. Appl. Sci. Res. 4, 282–286 (2008).
- Awwad, M. T. & Shbeeb, L. Am. J. Appl. Sci. 4, 390-396 (2007).
- Lucena, M. C. C., Soares, S. A. & Soares, J. B. Mater. Res. 7, 529–534 (2004).
- Emery, S. J. & O'Connell, J. in Proc. 7th Conf. Asphalt Pavements for Southern Africa 29 August-2 September 1999, Victoria Falls, Zimbabwe (CAPSA, 1999).
- Hofsink, W., Kong Kam Wa, N. Y. & Dickinson, M. N. in Proc. 8th Conf. on Asphalt Pavements for Southern Africa 12–16 September 2004, Sun City, South Africa (CAPSA, 2004).

- Hanyu, A., Ueno, S., Kasahara, A. & Saito, K. J. East. Asia Soc. Transport. Stud. 6, 1153–1167 (2005).
- 13. Fatehifar, E., Alizadeh Osalu, A., Kaynejad, M. A. & Elkamel, A. in Proc. 3rd IASME/WSEAS Int. Conf. Energy & Environment 23–25 February 2008, Univ. Cambridge, 330–335 (2008).
- 14. Lagzi, A. S. et al. in *Air Pollution Modelling and Simulation* (ed. Sportisse, B.) 264–273 (Springer, 2002).
- Aliahmadi, A. R., Jafari, M. & Amiri, B. in Proc. 2nd National Conf. Logistics & Supply Chain 20–21 November 2006, Tehran (2006).
- 16. Hallikas, J., Karvonen, I., Pulkkinen, U., Virolainen, V.-M. & Tuominen, M. *Int. J. Prod. Econ.* **90**, 47–58 (2004).
- 17. Ojala, M. & Hallikas, J. Int. J. Prod. Econ. **104**, 201–213 (2006).
- Harland, C., Brenchley, R. & Walker, H. J. Purchasing Supply Management 9, 51–62 (2003).
- The Experiences and Challenges of Science and Ethics: Proceedings of an American-Iranian Workshop (NAS, 2003).
 Allchin, D. Sci. Educ. 8, 1-12 (1999).

See Editorial, page 699.

Exoplanet claim bites the dust

Strike one planet from the list of 400-odd found around stars in other solar systems: a proposed planet near a star some 6 parsecs from Earth may not exist after all.

The finding is also a strike against a planet-seeking strategy called astrometry, which measures the side-to-side motion of a star on the sky to see whether any unseen bodies might be orbiting it. Ground-based astrometry has been used for more than a century, but none of the extrasolar planets it has detected has been verified in subsequent studies.

In May, Steven Pravdo of the Jet
Propulsion Laboratory in Pasadena,
California, and his colleagues raised
fresh hopes for the technique when they
announced an exoplanet, six times more
massive than Jupiter, orbiting VB10, a
star about one-thirteenth the mass of
the Sun, using a telescope at the Palomar
Observatory in southern California
(S. Pravdo and S. Shaklan Astrophys. J.
700, 623–632; 2009). But now a group
led by Jacob Bean at the Georg-August
University in Gottingen, Germany,
has used a different approach, and found
nothing. "The planet is not there," says Bean.

Bean and his colleagues used a well-honed technique called radial velocity, which has found most of the extrasolar planets detected so far. The method looks for shifts in the lines of a star's absorption spectrum to track its motion towards and away from Earth, which would be caused by the influence of a planet.

Radial-velocity measurements typically exploit the visible bands of the electromagnetic spectrum. But VB10 is a very dim star and gives off most of its light as infrared radiation. At the Very Large Telescope in Chile, Bean placed a gas cell filled with ammonia in the path of the starlight, enabling him to calibrate the instrument for the infrared.

"We would definitely have seen a significant amount of variation in our data if [the planet] was there," says Bean, who has submitted the work to the *Astrophysical Journal* (J. L. Bean *et al. Astrophys. J.* preprint at http://fr.arxiv.org/abs/0912.0003; 2009).



Pravdo says that Bean and his colleagues "may be correct, but there is hyperbole in their rejection of our candidate planet". Bean's paper, for instance, only rules out the presence of any planet that is at least three times more massive than Jupiter, says Pravdo, adding that the work "limits certain orbits for possible planets but not all planets".

"Unfortunately, astrometry is a very difficult business," counters Bean, explaining that Earth's atmosphere can introduce distortions that affect the measurements. Astrometrists rely on watching a field of stars about the same

distance away as the target star to calibrate their measurements, and that can be tricky, says Alessandro Sozzetti, an astrometry expert at the Turin Observatory in Italy. "Even if we think we have selected a good set of reference stars," he says, "we may still be limited by atmospheric effects that cause an extra jitter" in the motion of those stars.

Alan Boss, an exoplanet expert at the Carnegie Institution of Washington, agrees.

He points to the well-known 'detection' of 1963, when Dutch astronomer Piet van de Kamp used astrometry to claim that two planets were orbiting Barnard's Star — a finding disproved a decade later. The dispute over the VB10 planet, says Boss, "is another example of how hard it is to detect extrasolar planets using astrometry from the ground".

Astronomers expect astrometry to work much better above the distorting effects of the atmosphere. Two space missions in the works — the European Space Agency's GAIA, due to launch in 2012, and NASA's Space Interferometry

Mission, the launch date for which is yet to be set — will use the technique to search for planets as small as Earth around Sunlike stars, says Sozzetti. More significantly, astrometry can yield the mass of a planet, whereas radial velocity only puts a lower limit on it.

Bean admits that astronomers might one day find a planet around VB10 if they scrutinize the star long and hard enough. "The main lesson from VB10," says Boss, is that a lot of high-quality data are needed to be sure that an exoplanet is present.

Katharine Sanderson

Primate study halted by US university

Officials fear violent reprisals from a reinvigorated animal-rights movement.

Administrators at Oklahoma State University (OSU) in Stillwater have abruptly cancelled an anthrax vaccine study that would have killed dozens of baboons.

The project, funded by the US National Institutes of Health (NIH) and led by Shinichiro Kurosawa of Boston University School of Medicine in Massachusetts, had been approved by the OSU animal-care committee in September and was awaiting review by the biosafety committee when OSU president Burns Hargis vetoed it in October, calling the study "controversial".

Kurosawa had hoped to use the OSU animal facility because it has the required level of biosafety containment for anthrax. "As guest scientists at OSU, we are obliged to follow their policies, and it is unfortunate that we cannot fully complete our research there at this time," Kurosawa says. Along with collaborators at the Oklahoma Medical Research Foundation in



Burns Hargis has vetoed a primate study at Oklahoma State University.

Oklahoma City, the University of Oklahoma in Norman and the University of Chicago in Illinois, he planned to investigate the biochemical pathways that lead to death following anthrax infection, and to test an anthrax vaccine.

Some faculty members have suggested that the decision to cancel the study might be linked to pressure from Madeleine Pickens, the wife of oil magnate and OSU benefactor T. Boone Pickens. Madeleine Pickens had previously expressed disapproval of surgical training procedures involving animals in the university's veterinary school. Spokespeople for both Pickens and the university deny the suggestion. Hargis defended his decision in the Tulsa World newspaper, emphasizing that 124 animals could have been killed on campus.

"There are regrettably some violent acts committed by animal-rights groups," says OSU vice-president of research Stephen

McKeever, "and the president felt we should take our breath here and not do this project just yet." McKeever says the decision does not indicate a change in institutional policy, but that future proposals for primate studies will be considered in consultation with researchers, the OSU animal-care committee and his administrators. Indeed, for several

Culture clash at Australian synchrotron

The future of Australia's biggest scientific facility, the Australian Synchrotron in Melbourne, is in question. After its director was fired and its staff went on strike in recent weeks, relations between the board that oversees the facility and the scientists who use it have broken down

Australian and international researchers who sit on the synchrotron's Science Advisory Committee (SAC) are due to meet with board director and lawver Catherine Walter, and have threatened to resign unless Walter herself leaves, "If the committee resigns, no eminent scientist from around the world is going to touch the Australian Synchrotron with a bargepole," says a synchrotron researcher who asked to remain

"We have drawn a line in the sand," says committee chair Frank Larkins, of the University of Melbourne, "Of the 40-odd synchrotrons around the world, I don't think this has ever happened before." Larkins says that the row is partly a result of a culture clash between the business-oriented Walter and the scientists, "The 'bottom-line' priorities for the scientists are matters such as

publications in highimpact journals," he says, whereas the board is more focused on financial factors that "have to be balanced with the science priorities".

"What Australia hasn't realized is that it is trying to get an international scientific project to work," adds SAC member Michael Grunze of the University of Heidelberg in Germany. "It's not just a synchrotron issue. It's about trust, faith and the ability to manage big science facilities."

The synchrotron began operating in July 2007, and soon hosted more than 1,000 users on its nine beamlines. In the past 2 years, it has catered to scientists from almost 50 institutions in Australia and overseas, who use the synchrotron's intense light to study, for example, nanomaterials and proteins. Its

Aus\$221-million "No synchrotron (US\$201 million) cost director has ever was funded by an been removed unusual combination of the government of the state of Victoria, without notice." the Australian and

> New Zealand federal governments, universities and other institutions. It is governed by a board of seven people, including Walter.

According to scientists on the SAC, Walter has routinely ignored advice given by the committee, particularly about the need to begin securing funding for after 2012, savs Grunze, "In the first year we

never even got a reply recognizing that our recommendations had been received," he says. The SAC recommended in May 2008 that the synchrotron should begin working on a new 5-year funding plan that would raise Aus\$250 million. "Raising that much money is not easy," says Larkins. "You need to build a consensus on what the money should be spent on and what science you need to do. We advised the board accordingly, but the chair took no significant action." Walter says that the SAC "assists the board on scientific matters. It does not have governance responsibility."

The situation became critical in late October when the facility director, chemist Robert Lamb of the University of Melbourne, was dismissed without explanation, Lamb says. "We were doing extremely well," he says. "We were running just below budget, had raised a considerable amount of

from office



GOT A NEWS TIP?
Send any article ideas for Nature's News section to newstips@nature.com

faculty members, the biggest upset is not the decision itself but the fact that it was made without consulting them.

Veterinary researcher Richard Eberle, who was an administrative liaison for the study, believes that the affair might give the impression that the university is no longer a reliable research partner. He notes that two major proposals for OSU-based primate research, involving some of the same institutions, are pending at the NIH. The NIH Office of the Director said in a statement that institutions are expected "to complete NIH supported projects as requested, approved and funded".

The dispute comes during a time of heightened activity by animal-rights activists, including firebombings at two University of California campuses. Although few institutions have policies that prohibit primate research, not many are keen to establish new primate programmes, says Dario Ringach, a neurobiologist at the University of California, Los Angeles, who stopped working on primates because of pressure from activists. "It is changing the kind of work people will do in the future," he says. "If students come to me interested in primate research, I would tell them to think about other things."

Brendan Borrell
See Editorial, page 699.

money and were very popular with the national and international scientific user communities."

Walter says that legal and confidentiality issues prevent the board from disclosing the details of Lamb's sacking. "The board unanimously resolved this course of action for a number of reasons that included significant compliance and stakeholder relationship issues which had built up over the past year," she says. The synchrotron has secured operational funding until June 2012 and is working on plans for further expansion, she adds.

So far, neither the Victorian nor the national governments have intervened, saying that it is a matter for the board to resolve. Angry at this lack of action, the scientists and technicians at the synchrotron began limiting their work to 9 a.m. to 5 p.m. in late November, a crippling move for a facility that is booked to run 24 hours a day. Although they returned to normal schedules after Walter agreed to discuss their complaints, they now say that they will again work to rule unless Walter leaves.

The SAC has written to Victorian premier John Brumby protesting against Lamb's dismissal: "In our collective experience spanning decades and continents, no synchrotron director has ever been removed from office without notice."

Stephen Pincock

'Killer application' for protein synthesis is retracted

"This takes away one

of the benchmarks

people would cite

to show how far the

method could go."

The retraction of two papers from the lab of prominent US chemist Peter Schultz is a setback for researchers trying to synthesize and study glycoproteins — proteins with sugar chains attached.

The papers, published in Science¹ and the Journal of the American Chemical Society² (JACS), seemed to show that technology enabling the bacterium Escherichia coli to make proteins from many non-natural amino acids could also incorporate sugars at specific sites.

Schultz, of the Scripps Research Institute in La Jolla, California, says that while attempting to replicate the work in the two papers, members of his lab discovered that non-natural glycosylated amino acids — ones with attached sugars — behave differently from all other non-natural amino acids his lab has studied. The

researchers were unable to get the specific amino acids described in the two papers to integrate into proteins, although they did manage to get the bacteria to make proteins incorporating other glycosylated amino

acids, Schultz says. He and his colleagues retracted the *JACS* paper on 4 September³ and the *Science* paper on 27 November⁴.

In August, a paper co-authored by Eric Tippmann, a former postdoc of Schultz's who is now at Cardiff University, UK, argued that the method described in the papers could not have worked anyway⁵. *E. coli*, he reported, has insufficient levels of the relevant enzymes necessary to process the glycosylated amino acids that were used in the experiment. He and his colleagues suggest⁵ that the proteins reported in the retracted papers contained natural rather than non-natural glycosylated amino acids.

Schultz says it could be true that the proteins incorporated natural rather than non-natural amino acids, but adds that there are other possible explanations for his results.

He says that the conditions of the original experiments may have allowed the *E. coli* to process the glycosylated amino acids, which had been modified to allow them to enter the bacteria easily. However, the lab no longer has the notebooks detailing the original

experiments, so the team can't replicate those conditions, Schultz explains.

Schultz says that he had members of his lab try to replicate the papers for more than two years. "We worked hard on it, and there are real peculiarities associated with the glycosylated amino acids that we still don't understand," Schultz says. "We couldn't get it to work." Only then did the team decide to retract the papers. "I think we did the right thing," says Schultz.

Glycoproteins are ubiquitous in biology and pharmacology, but difficult to make artificially in living cell systems, so the ability to direct bacteria to make specific glycoproteins would have been a boon.

Chemist David Tirrell of the California Institute of Technology in Pasadena, who also studies methods for incorporating non-natural amino acids into proteins, says

> the retractions will be a blow for glycobiologists. But because the glycobiology work was often seen as proof of principle, it is also a disappointment for anyone working on making proteins from non-natural amino acids, he says. "This

takes away one of the benchmarks people would cite to show how far the method could go," says Tirrell.

Another former postdoc of Schultz's, Ryan Mehl, who is now at Franklin & Marshall College in Lancaster, Pennsylvania, agrees. "[Glycobiologists] went from something where they had the potential for great tools to zero, so it's a big deal for that field."

Schultz's underlying method for incorporating non-natural amino acids into proteins has been reproduced by other labs, note Tirrell, Mehl and other scientists. But the ability to incorporate glycoproteins "would have been a killer application", says Steven Benner of the Foundation for Applied Molecular Evolution in Gainesville, Florida. "I'm disappointed that it didn't work." Erika Check Hayden

- 1. Zhang, Z. et al. Science **303**, 371-373 (2004).
- 2. Xu, R. et al. J. Am. Chem. Soc. 126, 15654-15655 (2004).
- 3. Xu, R. et al. J. Am. Chem. Soc. 131, 13883 (2009).
- 4. Zhang, Z. et al. Science **326**, 1187 (2009).
- Antonczak, A. K., Simova, Z. & Tippmann, E. M. J. Biol. Chem. 284, 28795–28800 (2009).



Cattle disease faces total wipeout

Rinderpest goes the way of smallpox.

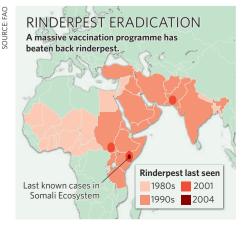
What does it take to wipe a scourge off the face of the Earth? A massive global push to hunt down and eradicate the last few stubborn pockets of disease — whether the problem is in people or cattle.

World health bodies say that within 18 months they will celebrate the eradication of rinderpest, the world's most devastating cattle disease. It would become only the second disease that humans have wiped from the globe — after smallpox, which was declared vanquished in 1980 — and will mark a "massive achievement for the veterinary community", says Chris Oura, head of the Non-Vesicular Disease Reference Laboratory Group at the Institute for Animal Health in Pirbright, UK.

"Rinderpest tops the list of killer [animal] diseases," says Juan Lubroth, chief veterinary officer for the Food and Agricultural Organization of the United Nations (FAO) in Rome. Just as smallpox ripped through human populations for centuries, so too has rinderpest drastically reduced animal populations.

Also known as cattle plague, rinderpest can lead to famine when people lose the beasts they need to plough their fields. It first spread from Asia to Europe in the herds of invading tribes, causing outbreaks in the Roman Empire in 376–386, and since then it has killed millions of cattle and other wildlife throughout Europe, Africa, the Middle East and the Indian subcontinent. The world's first veterinary science school was established in France in 1762 to train specialists to deal with rinderpest.

The disease, which can kill 80–90% of infected cattle within ten days, is caused by a morbillivirus — a group of viruses that also includes measles. Clinical signs include fever, discharges from the eyes and nose, diarrhoea and dehydration.





Vaccines developed in the 1980s have helped to control outbreaks of rinderpest around the world.

In the 1980s, outbreaks in Nigeria cost around US\$2 billion. But that decade also saw a breakthrough in controlling the disease: a vaccine containing the attenuated virus that was heat-stable and could be stored and transported over long distances.

Going global

In 1994, a global effort to eradicate rinderpest was launched, headed by the FAO and the World Organisation for Animal Health (OIE), based in Paris. It incorporated several earlier, regional efforts and focused on widespread vaccination programmes and on long-term monitoring of cattle and wildlife. The last known outbreak was in Kenya in 2001, with the last remaining pockets of the disease in Pakistan, Sudan and the Somali Ecosystem (parts of Somalia, Ethiopia and Kenya) thought to have been eradicated by 2007 (see map).

Oura says that the biggest scientific challenge in eradicating the virus is the large-scale monitoring and surveillance needed to ensure that the virus is gone. "It's a huge task when you have the virus in developing countries and war zones, such as Somalia, to carry out monitoring and surveillance," he says. By the 1970s, smallpox, too, was found only in the war-torn Horn of Africa, where the last case

was isolated in Somalia in 1977.

Although the rinderpest vaccine can provide life-long protection, it also poses a challenge. Because it contains the live virus, diagnostic tests can't differentiate between infected and vaccinated animals, as both will test positive for antibodies against the virus. Cows also pass on antibodies to their offspring through their milk. So, to confirm whether the virus has been eradicated, vaccinations must stop for a period of two years and calves younger than two years old then need to be tested. "It is a difficult, long process to make sure nothing is there," says Oura.

Lubroth says he is "confident" that the world is already free of the disease but that the FAO and the OIE expect to make an official declaration that it has been eradicated in 18 months.

Bernard Vallat, director-general of the OIE, says that the hold-up is because 12 countries are yet to submit their final test and surveillance results to the organization. Even after the disease is declared extinct in the wild, it will live on in the lab. Over the next year and a half, the OIE will be drawing up an inventory of which governments and laboratories around the world are keeping a stock of the virus for research purposes.

Natasha Gilbert



HAVE YOUR SAY Comment on any of our News stories, online. ww.nature.com/news

Centre turns away from healing herbs

US research hub on complementary and alternative medicine shifts towards symptom management.

A decade ago, US National Institutes of Health (NIH) director Harold Varmus was invited to quit when he opposed a senator's plans to elevate a small NIH office into a research centre for testing the validity of alternative therapies.

Having grown from those modest beginnings, the National Center for Complementary and Alternative Medicine (NCCAM) in Bethesda, Maryland, celebrates its tenth anniversary this week. It still draws fire from traditional scientists, who complain that by the very act of studying therapies outside the realm of conventional medicine, it has legitimized them.

But, many of the centre's clinical trials have ended up disproving alternative therapies rather than endorsing them. One study published in July 2005 (R. B. Turner et al. N. Engl. J. Med. 353, 341-348; 2005) showed that the herb echinacea neither prevented nor relieved the symptoms of rhinovirus infections, the most common cause of the 'common cold'; sales of echinacea fell markedly thereafter (see graphic). "We are in fact unbiased," says Josephine Briggs, who took over as centre director early last year, having overseen the NIH's kidney research for almost a decade until 2006. "We are directing research that will be rigorous."

In 2007, 38% of Americans said they had turned to alternative treatments at least once over the previous 12 months, spending US\$33.9 billion on a gamut of therapies from acupuncture to herbal remedies to yoga. Dwarfed by other research powerhouses on the NIH's Bethesda campus, the NCCAM this year spent \$125 million studying these treatments, or

about one-third of the NIH's total investment in alternative therapies; the rest is spent by other institutes and centres, particularly the National Cancer Institute.

Briggs and others say that the goal is to provide hard data on alternative therapies, regardless of whether that data debunks or affirms any given treatment. "One of [NCCAM's] critical roles is to actually weed out the snake oil, which I am sure there is quite a bit of," says Richard Davidson, a neuroscientist and NCCAM grantee at the University of Wisconsin, Madison. His work on the effects of meditation on the brain

and peripheral biology has been published in mainstream journals such as The Proceedings of the National Academy of Sciences, PLoS Biology and The Journal of Neuroscience.

Many US researchers still say such funding is a waste of time and money. "You are doing scientific research on treatment modalities that are not being used or promoted by science-based practitioners in the first place," says Steven Novella, a neurologist at Yale University in New Haven, Connecticut. "They never abandon a treatment because

the scientific evidence shows that it doesn't work. So what's the point?"

Briggs says that she would like to see the focus of the institute move away from overly "optimistic" attempts to cure chronic diseases, a trend she saw in grant applications after assuming the centre director job. "Investigators entered the project thinking a herbal compound might change diabetes or cancer," she says. "The main way I'm trying to adjust our portfolio is greater focus on symptom management, particularly greater focus on pain", for which many use alternative therapies.

Indeed, the list of "promising leads" flagged in the centre's draft strategic plan for the next five years is weighted towards pain management and mind-body medicine and silent on herbal therapies.

Briggs is also pushing an increased emphasis



Echinacea doesn't alleviate cold symptoms.

on basic and animal studies of the physiological mechanisms underlying alternative remedies. In its first decade, along edies. In its first decade, along with the echinacea work, the institute funded costly clinical trials studying whether St John's wort could relieve depression and, with the National Cancer Institute, whether vitamin E and selenium could prevent prostate cancer, among other studies. The results were resoundingly negative. Briggs concluded that the NCCAM needs to invest in fundamental work on natural products "before we invest more in clinical trials".

The centre "is increasingly defining a mission for itself that makes sense scientifically", says Bruce Rosen, an NCCAM grantee at Harvard Medical School in Boston, Massachusetts, who studies acupuncture's effect on brain function.

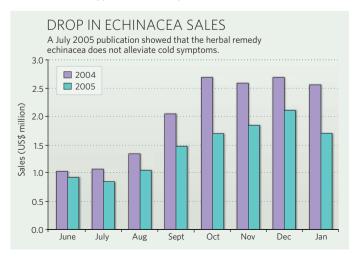
Briggs's approach has earned the respect of even the institute's fiercest critics. "I'm encouraged by Dr Briggs's receptiveness to comments and criticisms, and her commitment to altering the research priorities of the NCCAM," says Donald Marcus, an immunologist at Baylor College of Medicine in Houston, Texas. "My concern is the political constraints under which she operates."

Marcus and Novella are members of the Institute for Science in Medicine, a group of physicians and scientists newly set up to fight government policies promoting alternative medicine. He contends that the entire NCCAM

> enterprise "has been a remarkable waste of money", driven by Senator Tom Harkin (Democrat, Iowa), 🚊 who authored the 1998 legislation that established the centre. "The best thing they could do with the NCCAM is to dissolve it," Marcus says. "But that's not going to happen. Harkin's too powerful."

A decade after the NCCAM's birth, that power remains on display: Harkin and others have inserted provisions in healthreform bills in both houses of Congress mandating that insurers reimburse state-licensed alternative medicine providers.

Meredith Wadman



Israel weighs up new funding agency

Top scientists say that basic biomedical research is flagging.

JERUSALEM

Israel is a research powerhouse, sustained by a per-capita level of spending that, by some counts, is the highest in the world. But most of that money is spent by industry; in basic research, which is mostly conducted in highereducation institutions, Israel lags behind many other developed countries (see graph), with biomedical research bringing up the rear.

Israeli researchers are on a quest to remedy that by establishing the country's own biomedical research funding agency, along the lines of the US National Institutes of Health or Britain's Medical Research Council, and boosting funding by as much as tenfold.

At an international workshop last week in Jerusalem, leading researchers declared their support for establishing a National Fund for Biomedical Research (NFBR). The biggest hurdle remains the Israeli ministry of finance, where officials are reserving judgement on the proposal.

Compared with Israel's roughly US\$65 million in merit-based basic research funding each year, the United States, with 50 times the population, has 1,000 times as much federal funding for basic research. The US National Institutes of Health alone is a \$30-billion agency.

The Israel Science Foundation (ISF), which oversees government funding for basic research, has allocated just \$8 million annually for biomedicine in recent years. Another \$2 million in state support goes to clinical research through the ministry of health.

Immunologist Ruth Arnon of the Weizmann Institute of Science in Rehovot, who is also vicepresident of the Israel Academy of Science and Humanities, leads a task force that last year published a report calling for the establishment of the NFBR. It concluded

that Israel significantly underutilizes its potential resources in the biomedical research field owing to limited research funds, outdated infrastructure and the difficulty that young researchers have in finding jobs in Israel.

Ruth Arnon calls for a boost

to biomedical research.

"Significantly expanding biomedical research in Israel can bring Israeli scientists back from overseas, create a healthier society - and ensure that Israel is known throughout the world for its science rather than for

ISRAEL'S RELATIVE R&D SPEND Industry Higher education 100 3ERD/GERD (%) HERD/GERD (%) 80 60 40 10 GERD, gross domestic expenditure on R&D Switzerland Israel Sweden BERD, business expenditure on R&D United Kingdom United States HERD, higher-education expenditure on R&D

violence and conflict," she says.

A new national fund for biomedical research should aim to give out \$100 million annually, she says. She predicts that the NFBR could be up and running by 2011.

At the workshop last week, Daniel Hershkowitz, the minister for science and technology, said that next year he would bring the proposal before the ministerial committee on science and technology, which has the power to make policy decisions on behalf of the government. Deputy health minister Yakov Litzman also pledged support for an

> increase in funding, and Meir Sheetrit, chairman of the science and technology committee of the Knesset, Israel's parliament, said that he will bring the proposal up for discussion there.

But a finance-ministry official attending the workshop, who asked not to be named because he was not authorized to comment, was less sanguine. He said that he still needed to be convinced that new money for biomedical research should take precedence over research funding in other fields. He also noted that, statutorily, his ministry

cannot allocate money for a specific research purpose; research funds are allocated to the Council of Higher Education, an independent body that accredits and funds Israeli universities. The council then decides how much of its budget to hand over to the ISF.

Others have more creative ideas about where the money might come from. Sheetrit suggests it could be diverted from the \$350-million budget of the chief scientist in the ministry of industry, trade and labour, which is the government's main office for funding applied research.

To ease the financial strain, Arnon suggests that the NFBR could be established modestly, with an initial budget of \$20 million, and gradually ramp up funding. It could, she says, start out as a division of the ISF before becoming a fully fledged independent funding body.

Arnon thinks the NFBR could remedy the flaws she sees in the current system. For one, ISF grants are small, limited to a maximum of \$50,000 a year for five years, whereas a modern biomedical research laboratory usually costs at least three times that much to run. $\frac{1}{2}$ multiple grants.

"The average Israeli scientist spends a third of his time writing grants and grant reports," Arnon says. The NFBR would allow researchers to receive adequate funding from a single source, she says, enabling them to devote more time to their scientific work.

Arnon also singles out clinical research as in need of a fix. "We interviewed quite a few clinicians in preparing our report, and they said that the most critical factor preventing them from performing research is a lack of protected time," she says. "Clinicians come to work early in the morning and work late into the afternoon and their time is completely consumed by clinical tasks."

Undeterred by the coolness of the financeministry official last week, Arnon says that she and her colleagues are talking to policy-makers at the finance and other ministries. "The money is there — it's a matter of deciding priorities," she says. "I'll be knocking on doors and making my case. We have to do this."

Haim Watzman





Copenhagen: the scientists' view

The United Nations Climate Change Conference is mainly a political affair but it has drawn hundreds of scientists to the Danish capital. **Jeff Tollefson** finds out what they hope to gain.

s the United Nations summit on global warming kicks into gear in Copenhagen this week, upwards of 15,000 people are converging on the city. The official negotiators from 193 countries will spend much of their time behind closed doors at the Bella conference centre, but they will be a minority of the visitors. Orbiting around the negotiators will be representatives of almost every segment of society, including hundreds of scientists.

The researchers will attend scheduled science sessions and gather for countless impromptu discussions in corridors and cafeterias. Many are presenting their latest work — on a vast array of topics including forest carbon, emissions scenarios and green technologies. Some hope to influence policymakers and provide technical advice on issues that emerge during the negotiations. Others are coming to educate themselves about the treaty process and to network.

A climate summit is a flurry of activity, with the central negotiations surrounded by side shows that last from early in the morning until late at night. When the formal sessions finally wind down (if, in fact, they do, as negotiations have been known to go all night), discussions often continue over dinner and drinks.

The Copenhagen meeting, which runs from 7 to 18 December, is officially the 15th Conference of the Parties (COP) to the United Nations Framework Convention on Climate Change, which was signed at the Earth Summit in Rio de Janeiro in 1992. Negotiators have been meeting each year for a COP since 1995, but the expectations and the stakes for this summit are orders of magnitude higher than for any previous one.

Twelve years after taking their first tentative steps with the Kyoto Protocol, countries are now aiming to restructure the global economy and to lock in deep cuts in greenhouse-gas emissions for decades to come.

In advance of the summit, *Nature* talked to researchers from around the world about how they plan to take part.

Ottmar Edenhofer, co-chair, Working Group III, IPCC; deputy director of the **Potsdam Institute for Climate Impact** Research, Germany.



"I wouldn't say that I am depressed, but I feel very sad about the negotiation process as it stands now. But I don't see that this can be changed substantially by scientists."

Edenhofer is wearing two hats in Copenhagen. As co-chair of the Intergovernmental Panel on Climate Change (IPCC) working group on mitigation, he is presenting results from the group's 2007 assessment at several side events and briefings for policy-makers. But those events frequently lead to additional contacts, requests and conversations in which, as an independent scientist, he can offer his own thoughts on the latest research and what it means for policy-makers.

Edenhofer says that the negotiations are falling short of what is needed to address global warming and that scientists are unlikely to change that now. However, he argues it would be wrong to downplay the role of science in the process. Scientists were the first to raise concerns about climate change, and the IPCC's fourth assessment has served as the foundation for the negotiations. He sums up the IPCC's findings this way: humans cause climate change; climate change has severe impacts; and it is not too costly to reduce emissions. "These three messages have already changed the mindset of the negotiators," he says.

Beth Sawin, biologist and programme director, the Sustainability Institute, Hartland, Vermont.



"We have this philosophy that 💆 if science is going to be helpful, if science is going to be helpful, \geq it has to show up, wanting to serve. What can we do to our model to make it more useful to somebody who is incredibly busy, overwhelmed, with not enough time and a huge responsibility?"

The Sustainability Institute has developed user-friendly climate-modelling software that can be run on a laptop computer to help negotiators assess the ultimate impact of any given emissions scenario. Negotiators can manually adjust the emissions and other parameters to analyse their own proposals as well as those of other countries; the model spits out forecasts for variables such as future temperatures and sea-level rise. In Copenhagen, Sawin says, the team is providing a "widget" that can be installed on computers to get the latest climate readings whenever Sawin's group updates its model with any new commitments announced by countries.

The application has generally received positive feedback from negotiators, but Sawin acknowledges the sobering reality that some delegates are less interested in detailed climate projections than in the next election in their home country. Nonetheless, she finds the whole affair touching. "I see that there are warts, and there is unfairness, and there are flaws in this process, but at least it's happening," she says. "So when I come home and talk to my kids, that's what I emphasize: that we happen to be alive at a time when people are trying to make common decisions about how to protect our common planet."



Albert Binger, science adviser to Grenada and the Alliance of Small Island States.



"I never had the slightest notion in my mind that one day I would be the guy telling everybody that the [target of] 2 °C the majority of the world wants is absolutely crazy. 2 °C is too much for too many people."

Raised in the mountains of Jamaica, Binger did a brief stint as a chemical engineer in the petroleum industry before earning a doctorate in agronomy at the University of Georgia. Today he is an official delegate advising island nations that are seeking to limit average global warming to 1.5 °C — or preferably less. Regularly oscillating between anger and a healthy island humour, he says. "Everybody needs to clean up their own goddamn mess."

Although Binger has full access to the talks, he leaves negotiating to the negotiators. His job is to harness scientific evidence in the push for more stringent greenhouse-gas targets. In practice, this means helping to answer questions that arise during the talks and providing scientific evidence for use in speeches and debates. As an islander who stands to lose everything to ocean acidification and sea-level rise, Binger takes the issue personally. "We want 1.5 °C or less, and we don't really ask it selfishly. Every person on this planet is better off at 1.5 °C than they are at 2 °C. I can sleep very easily with that."

Lawrence Buja, climate modeller, National Center for Atmospheric Research (NCAR), **Boulder, Colorado.**



"To a certain degree, the physical modellers have a much easier job than these politicians. Our molecules don't think for themselves and start doing different things midstream."

At COP 14 in Poland in 2008, Buja gave a briefing on NCAR's climate-modelling results for the fourth assessment of the IPCC, issued in 2007. He headed the modelling team at the time but is now directing a new group that is developing integrated climate models that include social and economic forces. His career change reflects a larger shift — Buja goes so far as to call it a "sea change" — for NCAR as an institution. Physical modelling will remain a core activity as scientists seek to clarify and provide more detail about the potential impacts of greenhouse gases, he says, but NCAR recognizes that it needs to provide policy-makers with more information about potential solutions.

In Copenhagen, one of his colleagues is presenting modelling results analysing the level and timing of emission-reduction targets, focusing on the 2050-2100 time frame. Buja is on hand to talk about these issues as well as to answer questions about the physical modelling, which is now being ramped up for the IPCC's fifth assessment, due out in 2014. But information flows both ways at these meetings, he says. "What this exposes the scientists to is how these negotiations and agreements are developed and what our role in informing them might be."

Martin Parry, climate scientist, Imperial College London.



"For individual scientists like me, frankly, many would say there's not much point in going. But I think it's a chance to meet those at the fringes of the political system who potentially do have quite a lot of leverage."

Were you to bump into him in Copenhagen and ask how the negotiations are going, Parry says he wouldn't have a clue. He has minimal or no contact with negotiators but says he finds value in exchanging ideas with scientists and activists. Those discussions can be particularly important, Parry says, because advocacy groups such as the WWF can then inject the latest scientific thinking into the political process as they lobby negotiators and government officials. In Copenhagen, he is expecting to participate in two side events, one on development issues and a second on agriculture.

Parry is also thinking about how to assess a major hole in how the world intends to respond to climate change. Some impacts can be avoided ≥ by reducing greenhouse-gas emissions. Others can be managed with enough money; in the vulnerable developing world, that means financial aid from wealthier nations. But the current proposals for emissions cuts and mon-etary support are not enough to avoid major impacts. "We're trying to close a gap here." impacts. "We're trying to close a gap here, coming at it from both ends," he says. Parry hopes that framing the issue this way — and quantifying the impacts — in Copenhagen will clarify where the policy-makers are coming up short, both in terms of emissions reductions and money for adaptation.

Paulo Moutinho, research coordinator, **Amazon Environmental Research** Institute, Brasilia, Brazil.

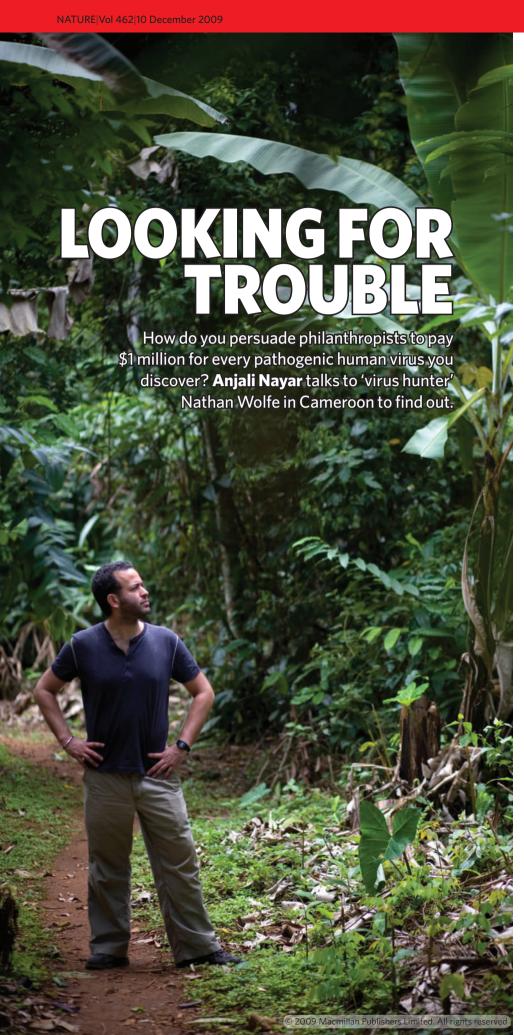


"I believe that [the forest- \opinion" protection strategy called] REDD could make a difference in COP 15, not just as $\frac{1}{2}$ a way to address emissions 5 from tropical deforestation, but also to create a new kind of synergy among nations. I believe that. That's exactly why I am going."

Moutinho started his career studying ants but has spent most of his time in recent years looking at ways to use carbon markets to stem emissions from deforestation while protecting biodiversity and the rights of indigenous people. Hopes have faded for a complete treaty in Copenhagen, but he is holding out for a significant decision on the forest-carbon component known as Reducing Emissions from Deforestation and Forest Degradation (REDD). Through REDD, wealthier nations seeking to 'reduce' their emissions would provide money for developing nations to protect their forests. In Copenhagen, Moutinho is presenting his organization's latest work on REDD in the Amazon and discussing Brazil's national greenhouse-gas commitments.

For him, Copenhagen is a perfect fit. Spending time in the field and publishing papers in Nature or Science is one thing, he says, but the goal must be to translate results into a digestible form for policy-makers. "Science is a tool to reach sustainable development. That's my view about science, and that's exactly what I'm doing." And when it comes to REDD, Moutinho says, the science is evolving rapidly and still plays an important part in the negotiations. Jeff Tollefson covers climate, energy and the environment for Nature.

For more on Copenhagen, see www.nature.com/ roadtocopenhagen.



very day, more than 100 patients line up for treatment outside the bare cement walls of a rural health clinic in the Niete forest of southern Cameroon. Most of them suffer from what virologist Nathan Wolfe calls "the usual suspects": malaria and typhoid. But every once in a while there is something a little different: a case that is tough to diagnose, or that doesn't respond as expected to medication. For the patients and the medics here, this presents a problem. For Wolfe, it is also an opportunity. "I get all excited," he says. "These areas are choc-a-block full of interesting, unusual viruses."

In 2005, Wolfe and his colleagues identified two unusual viruses that had jumped from non-human primates into Cameroonian hunters¹, and that are closely related to pathogenic viruses that have spread through the world's population. Since then, Wolfe has made a name and a livelihood for himself in search of more. In 2007, he founded the Global Viral Forecasting Initiative (GVFI), an effort to monitor for new viruses and other microbes in communities around the world that are in intimate contact with animals. By identifying these agents as soon as they leap into humans, and before they start passing easily between them, Wolfe believes that researchers can predict and prevent human pandemics before they start, averting the next HIV, SARS, Ebola or pandemic influenza, diseases that have all been traced to animals.

Brand Wolfe

Other researchers share that belief and are involved in this type of infectious disease surveillance. But what sets Wolfe apart is his swashbuckling style — he chooses to do most of his work in the field — combined with a flair for communication and negotiation. "He's in the minority of scientists; he's good at science, politics, media and PR," says Jeremy Alberga, the chief operating officer at the GVFI. "We have a brand and it's Nathan Wolfe and the GVFI," he says, "not the GVFI and Nathan Wolfe."

That brand has sold well. In 2008, Google. org, the philanthropic arm of Google, and the Skoll Foundation, an organization in Palo Alto, California, that supports social entrepreneurs, announced they would invest up to US\$11 million to help expand the GVFI's work in Africa and southeast Asia. In October 2009, the United States Agency for International Development (USAID) named the GVFI as a main partner in its \$330-million Emerging Pandemic Threats programme, a project involving various experts in wildlife surveillance, and for which the GVFI will receive tens of millions of dollars over the next five years. Last month,



Wolfe unveiled a new \$600,000 update to his lab in Yaoundé, Cameroon's capital, marking ten years of research in the country.

Wolfe and his work have also found a wider audience, featuring in news outlets ranging from Wired to Men's Journal. Earlier this year, Rolling Stone magazine ranked Wolfe at number 53 of "people who are changing America", and an article in Popular Science magazine was simply titled: "Nathan Wolfe: did we mention this guy was brilliant?" "We joke about it sometimes," says Frank Rijsberman, a programme director at Google.org, about the unrelenting press Wolfe receives. "He's definitely audacious — he has big visions — but he has enough of a track record in the field to do it realistically."

Lasting mark

Other 'virus hunters' say that they are less interested by Wolfe's semi-celebrity status, and more interested in what he is going to

"The chimps were up

to their eyeballs in the

blood and body fluids

of all these animals."

— NathanWolfe

find in the future. "Nathan's a bright, energetic young guy and I anticipate he will be successful," says Ian Lipkin, a specialist in pathogen surveillance and discovery at Columbia University in New York, who recently started collaborating with Wolfe.

"For the sake of the field it's important that he is. A lot of resources are allocated to Nathan. He's going to have to deliver."

"The real breakthroughs are published in the literature, and not on CNN," adds Peter Daszak, a disease ecologist and president of the Wildlife Trust in New York, who is working with Wolfe on the USAID initiative (and says he has been on CNN, a US cable news network, himself). "If you look back on this field in 10 years, you will say who actually did the discovery, was



Nathan Wolfe, above centre, is trying to find viruses that jump species as a result of bush meat hunting.

there a paper that came out of this group and did they prevent the next disease."

Wolfe's interest in virus emergence dates back to his master's degree in biological anthropology in the 1990s. Working under Marc Hauser and Richard Wrangham at Harvard University in Cambridge, Massachusetts, Wolfe was studying the behaviour of wild chimpanzees to see if they were consuming certain plants medicinally. He never answered the question — but while tracking chimpanzees in the forests of western Uganda, Wolfe witnessed several dramatic ambush hunts. "The chimps were up to their eyeballs in the blood and body fluids of all these animals," he says. "I just thought to myself what a perfect scenario for cross-species transmission of microbes." He postulated that the rural human populations of Central Africa, with a diet heavily dependent on bush meat, could be affected the same way. "I just thought it would be this wonderful grab-bag of new things jumping into humans and bouncing around," he says.

Wolfe started testing his ideas, looking at cross-species viral transmission first during doctoral work in Malaysian Borneo, and then back in Central Africa after being recruited by Donald Burke, an infectious-disease expert at

> Johns Hopkins University in Baltimore, Maryland. Wolfe started to follow subsistence hunters into the forest as they caught and butchered monkeys and other wild animals. As the hunters walked back with the catch over their shoulders, blood would run

down their backs and into the cuts on their bare legs and feet. Wolfe thought that there was a high chance that viruses were passing between them, and he collected blood from the hunters and screened for retroviruses related to those found in non-human primates. There was precedent: researchers think that the retrovirus HIV was originally transmitted to humans from non-human primates in Central Africa, perhaps during hunting or butchering.

In 2004, Wolfe showed that around 1% of

those he tested had been exposed to a retrovithose he tested had been exposed to a retrovirus called simian foamy virus, and sequence analysis showed that the virus had been passed into humans in the past from a mandrill, a gorilla and a type of monkey known as De Brazza's guenon². The next year he reported the discovery of two new human retroviruses — called human T-lymphotropic viruses (HTLVs) — one of which is related to a group of simian viruses and both of which are likely to have crossed over from non-human primates¹. Another member of this viral group, HTLV1, is also thought to have spread into humans and is associated with leukaemia and other conditions. Wolfe's studies suggested that certain viruses were passing from non-human primates quite freely and frequently into humans. "Retroviruses are jumping over all the time," says Wolfe, even though "only a few of them end up being significant".

Wolfe's work convinced him that the significant viruses could be identified and eliminated with proper surveillance of areas where crossspecies transmission was a particularly high risk. Much of the challenge though, as other researchers have found, lies in getting funding to do the kind of large-scale field studies and genetic sequencing that this surveillance requires. Microbe discovery "doesn't rate real well" in funding applications that tend to focus on known threats, says Gregory Gray, who studies zoonotic infections at the University of Florida, Gainesville. It's hard to justify science "when you don't even know exactly what you are going to encounter", he says.

Wolfe was helped by his timing, his scientific findings and his eloquence. In 2005, he secured one of the National Institutes of Health five-year, \$2.5-million Pioneer Awards, and the next year a tenured position at the University of California, Los Angeles (UCLA). Soon he was building up a public profile, with speaking arrangements around the world. Interest in wildlife surveillance had been rising on the back of scares over West Nile virus, SARS and the H5N1 avian influenza. "Nathan has been riding the wave of interest in the field," says William

Karesh, director of the Wildlife Conservation Society's Global Health Program. "It's like surfing, and the wave started getting really big," he says. "And, you know, if you are on a big wave, you look better if you know how to surf well."

Wolfe caught the eye of Google.org, which was looking to invest money and information-technology expertise in the field of emerging infectious diseases. The fact that Wolfe is "media-genic", was key, says Rijsberman. The parties discussed the collaboration for nearly a year before the \$11-million funding deal was signed in October 2008. Wolfe had founded the GVFI the previous year, but the agreement "made it real" says Wolfe, and he left his position at UCLA to head the institute full time.

Viral marketing

Part of Wolfe's magnetism for money and media seems to lie in the niche he has carved out for himself as an adventuring field virologist — going into the field to collect samples himself, rather than sending students or postdocs to do it. This also helps the science. Lipkin says that tissue or blood samples can easily become contaminated — plus samples are best frozen quickly. "It has been difficult to collect high-quality materials and move them

from remote sites in the developing world to laboratories," says Lipkin. "Nathan is doing us an enormous service with these collections."

Wolfe thrives on the difficulties involved in securing the right samples from people or animals. "There is a lot of logistics in what I do, there's a political angle to what I do, and there's a tremendous amount of negotiation," he says. "I love when people say: you simply can't collect specimens like that in China, or it is impossible to have a long-term program in the Democratic Republic of the Congo. I sort of enjoy that challenge."

The GVFI now runs human and animal collection sites in the Democratic Republic of the Congo, Malaysia and China — some of these from the types of markets where humans and animals are cheek by jowl, sites that helped SARS and H5N1 emerge and spread. Wolfe narrows down his hunt by focusing on these 'hot spots' for emerging disease, along with pathogens (such as retroviruses) that have a history of making the jump. "It's not like you are picking randomly from space what's going to be your next pandemic," says Wolfe, who maintains academic affiliations with Stanford

University in Palo Alto, California, and with Johns Hopkins University.

The Cameroon team is more than 30 strong, and has collected blood and tissue samples from more than 27,000 humans and 26,000 animals. With the re-vamped laboratory there, equipped with PCR machines to amplify DNA, and separate facilities for animal and human work to avoid contamination, Wolfe plans to boost the group's capacity to do molecular

"If you are on a big wave,

you look better if you

know how to surf well."

William Karesh

analyses on-site. But for the most part, he collaborates with world-class laboratories in Europe and the United States for his genesequencing work. Identification of a new virus is only the first step though: to spawn a

pandemic, a virus has to adapt to humans and spread easily between them, something that can be examined by surveillance and modelling. Wolfe is monitoring the novel viruses he identified in bush-meat hunters to see if they are spreading.

How to stop a potential pandemic, if Wolfe finds one, will depend on the case in hand, but might involve some combination of diagnostics, behaviour change and vaccines or

HUNTED
NHRS
READY FOR
CHIPPING
NHRS
REPOY
FOR
CHIPPING

SHIPPING

The GVFI collects thousands of blood samples from non-human primates (NHPs).

treatments, if they could be found. Karesh points to the Republic of the Congo, where scientists found out that Ebola outbreaks were spread into humans through handling dead gorillas and chimpanzees. Communities were trained to avoid handling sick animals and there has barely been another outbreak since. Wolfe's group is already working on prevention programmes that might reduce transmission — such as encouraging hunters to wrap slaughtered animals in leaves when they are being transported.

Wolfe has not yet identified a nascent pandemic, but he is finding out something about

existing ones. Earlier this year, his team used samples from chimpanzees in Cameroon and the Côte d'Ivoire to shed some light on the origin of the human malarial parasite *Plasmodium falciparum*. Genetic analyses showed that all existing populations of *P. falciparum* originated from the chimp version, *Plasmodium reichenowi*, and that the chimp parasite probably transferred across species between 10,000 years and 2–3 million years ago³.

There is pressure on Wolfe to deliver much more. Only \$5 million of the \$11 million from Google.org and the Skoll Foundation came up front and the group is eligible for another \$1 million every six months — if they

earn it. They earned the first \$1 million by collecting nearly 10,000 new samples and enrolling around 900 people in their studies in Asia. Rijsberman, who now sits on the GVFI board, helped the team to establish goals they call "scientific home runs": the origin of a major infectious disease or the identification of a new pathogenic virus circulating in human populations. With each home run, the group earns another \$1 million. The group has submitted

the malaria work to the Google. org board and hopes to hear in the next few weeks if it qualifies as their first home run.

Wolfe, aware of the scientific competition, is careful not to say much about what his next million will be for. But he surely has a strategy to earn it.

During a retreat to the coastal Cameroonian resort town of Kribi last month, Wolfe and his team were plotting their next scientific articles, making notes on papers and laptops. After a day of hard work, they sit down for drinks and take in the setting sun. Wolfe takes on his laboratory director in a

game of backgammon, starting the game with an aggressive move. "Nathan always has to do things differently," says Alberga, looking on.

After a few more dice rolls, it's neck and neck. But, despite his nonchalance, Wolfe is constantly calculating probabilities — and in the end, he comes out on top.

Anjali Nayar is an International Development Research Centre fellow at *Nature*.

- Wolfe, N. D. et al. Proc. Natl Acad. Sci. USA 102, 7994-7999 (2005).
- Wolfe, N. D. et al. Lancet 363, 932-937 (2004).
- 3. Rich, S. M. et al. Proc. Natl Acad. Sci. USA **106**, 14902-14907

CORRESPONDENCE

Emissions: Canada should not be isolating itself

It should be possible to draw up a fair, ambitious and legally binding treaty on climate change after this week's United Nations summit in Copenhagen (see go.nature.com/sRCuKV). But the next few years are going to be crucial if we are to stand a reasonable chance of keeping global warming to below an average of 2 °C.

Pessimistic media reports reflect the views of only a few world leaders, including Canada's, who lack ambition and imagination. Other leaders, including the US and Chinese presidents and the summit host, Prime Minister Lars Løkke Rasmussen of Denmark, are still pushing for a strong agreement.

The science is clear. To limit average global warming to the 2 °C agreed by world leaders at July's G20 meeting, total greenhouse-gas emissions will have to peak around 2015 and then decline sharply. If emissions are more than 25% above 2000 levels in 2020, the risk of exceeding 2 °C in this century would be more than 50%, even if emissions were reduced to low levels by 2050 (M. Meinshausen et al. Nature 458, 1158-1162; 2009).

Canada should be playing a strong multilateralist part, fulfilling its international commitments, offering constructive proposals on difficult and tense issues, and looking for compromise when positions are entrenched. But Canada has not been constructive and it has isolated itself. It is the only country to have signed up for a target under the Kyoto Protocol and then stated that it has no intention of meeting it. Canada has put forward positions that have heightened, rather than eased, tension. It has argued that developing countries should take on 'hard caps' (which emissions must not exceed, even if production increases) — something no other country advocates. And

it is uncompromising on issues such as using 1990 as a base year, which has blocked agreement.

The Copenhagen summit is a big challenge because industrialized countries such as Canada have yet to accept fully that they have a greater responsibility than developing countries for the problem of climate change, and a greater capacity to solve it. Because rich countries have used up so much atmospheric space in the course of their own development, they need to deliver substantial support so that poorer countries can develop without following our dirty path. Wealthy countries need to provide clean technologies and financing some US\$160 billion per year — so that developing countries can curb their own emissions and adapt to climate changes that are already happening.

Unfortunately, Canada is among the rich countries that have impeded progress in negotiations, and not just because it has disavowed its commitments under the Kyoto Protocol. Its 2020 emissions-reduction target is the weakest of all Kyoto parties' and it has not supported any options to deliver financing to the developing world. The needs are great. But so are the responsibilities and the opportunities. Canada and other wealthy nations can and must do

David Suzuki David Suzuki Foundation, Suite 219, 2211 West 4th Avenue, Vancouver, British Columbia, Canada V6K 4S2 e-mail: contact@davidsuzuki.org

Emissions: taking a collaborative lead will work better

The developing world is urging rich countries to take the lead in tackling climate change, as Jiahua Pan points out in his Opinion article (*Nature* **461**, 1055; 2009). However, in doing so, they should not overlook opportunities for technological

collaboration as a means to work towards achieving a global low-carbon economy.

Within the past few weeks, both the United States and China have announced targets to curb their carbon emissions. With the climate-change summit in Copenhagen this week, the world is expecting further dialogue and collaboration on climate change between the two biggest carbon emitters. Developed nations will benefit, in the long run, from technological cooperation with developing countries.

Many carbon-emission reduction technologies, such as carbon capture and storage, are still at an experimental stage, with the United States, Germany and Japan leading their research and development. These countries should expand the scope of their experiments to test commercial feasibility. Many new power plants will be constructed in developing nations, offering a good opportunity for cooperative projects with developed countries to apply new technologies, if agreements can be reached on clean-energy and emissionsreduction programmes.

Such fruitful collaborations on technical projects will increase the chances of successful implementation of practical approaches in addressing climate change.

Xufeng Zhu Zhou Enlai School of Government, Nankai University, 94 Weijin Road, Tianjin, 300071, China e-mail: zhuxufeng@nankai.edu.cn

For more about the Copenhagen conference, see page 714 and go.nature.com/sRCuKV.

Need for religions to promote values of conservation

Boris Hillmann and Jan Barkmann suggest in their Correspondence that the world's religions could make a positive contribution to biodiversity conservation (*Nature* **461**, 37: 2009). This

was also a major theme in the interfaith Parliament of the World's Religions, which as held in Melbourne, Australia, this month (www.parliamentofreligions.org).

But the attitudes of different religions towards conservation can vary greatly. Compare, for example, the anthropocentric mainstream Muslim position that humanity has responsibility for Allah's creation, which has been given to us as a gift, with the biocentric Jain belief that every being — animal, plant or human — has a soul and should be treated with respect.

Religious practitioners are more likely to help fund the protection of species and habitats if their religion formally values biodiversity. Until recently, most faiths had not formulated a stance on conservation practice, as few religious traditions explicitly promote it. Networks such as the Forum for Religion and Ecology at Yale University (http://fore.research.yale.edu) have promoted discourse by searching for environmental inspiration in religious teachings.

It is not enough to rely on the idea that sacred places may act as de facto biodiversity reserves, because sacred sites can still be overexploited. Embedding some of the values of conservation biology into the teachings and practice of the four billion people affiliated with one of the world's 11 mainstream faiths is a promising way forward.

Matthew Hall Centre for Middle Eastern Plants, Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK email: m.hall@rbge.ac.uk John Grim, Mary Evelyn Tucker Yale University School of Forestry and Environmental Studies, New Haven, Connecticut 06511, USA

Contributions may be submitted to correspondence@nature. com. Published contributions are edited. Please see the Guide to Authors at go.nature.com/cMCHno. Comments and debate are also welcome at our blog Nautilus: http://blogs.nature.com/nautilus.

Vol 462|10 December 2009

OPINION

Time to future-proof plants in storage

Seed banks must collect and condition their holdings of wild species so that they can thrive in landscapes transformed by climate change, say **Jeffrey Walck** and **Kingsley Dixon**.

hen the Millennium Seed Bank Partnership (MSBP) of the Royal Botanic Gardens in Kew added its 24,200th species to its collection this October — a Chinese pink banana — it completed one of the major tasks set at its inception in 2000: to store seeds from 10% of the world's known wild flora. Its next target is even more ambitious: to bank a quarter of all wild plant species by 2020. Holding seeds against the immediate threat of extinction is crucial to preserving the world's plant diversity. Yet storage is only part of the challenge. Seed banks will have to do much more to re-establish species in landscapes that, because of global warming, will have changed irrevocably since they were collected.

The MSBP and other seed banks such as the Svalbard Global Seed Vault in the Arctic. nicknamed the 'doomsday vault', endeavour to collect samples that represent the normal genetic diversity of a species. At low temperatures, seeds can remain viable for hundreds if not thousands of years. Herein lies the problem: such seeds are literally frozen in time, a snapshot of the genetic diversity of a species at a particular point. Attempts to revive a seed in future habitats very different from those in which it developed could be doomed to failure. A germinating seed is genetically programmed to respond to a precise interplay of temperature and moisture that determines the climatic conditions in which it can best develop. In an environment that lacks the right temperature and moisture balance — which is likely to be disrupted by climate change — a seed will either fail to germinate or the seedling will perish soon afterwards (see Fig. 1).

Seed banks are expert at the preservation of wild seed. To prevent their samples becoming museum pieces they also need to take responsibility for the way the seeds are collected and

for preparing them for planting back in the wild. Shrewd collection strategies can help to ensure that species are better adapted for replanting in habitats likely to differ substantially from their

natural ranges. The primary aim should be to harvest as much genetic diversity as possible, so that collections are more likely to contain plants with traits suited to a changed climate.

One way to do this is to collect species at the limits of their ranges, where plants often

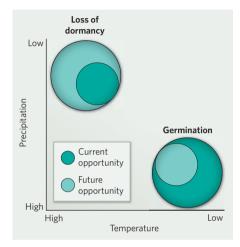


Figure 1 | Changes in climate could affect the success of stored seeds. For seeds that need warm, dry conditions for dormancy to end, and cool, moist conditions to germinate — such as Western Australia's kangaroo paw — future warming and drying may enhance dormancy loss, but offer a smaller window for germination.

have higher levels of genetic novelty — for example, Mediterranean floras at the arid margins of their habitats, or mountain floras at the upper and lower limits of their altitude ranges. Another strategy is to collect samples from the same locations at different points in time, say every 10 or 20 years, because, as the climate changes, wild populations will produce natural variants that are better adapted to the new conditions.

Mixing varieties

"Harvest as much

genetic diversity

as possible."

Once gathered, much can be done to condition seeds for reintroduction to the wild. Most restoration nurseries propagate plants from stored seeds under glasshouse conditions without the robust selection processes of natural

conditions, so it will be necessary to manipulate the growing environment and to screen for plants most suited to the new environment. There are established methods for doing this.

For instance, subjecting seedlings to increased temperatures could allow the selection of those with higher heat tolerance. Likewise, subjecting seedlings to dry conditions could increase drought tolerance in subsequent generations. Both approaches could require multiple

regeneration cycles at significant cost.

When the time comes to replant wild habitats, restoration scientists can use several techniques to maximize the chances that seeds will thrive in an altered landscape. For example, models are available that will predict how climate change will alter the home range of a species, so restorers can be sure to reintroduce seeds in places that will match their particular germination requirements. Where a plant population is heading for extinction and there are no stored samples suitable for repopulating, one option is to mix seeds from across the species' natural range — for example in a mountain habitat, mix seeds from higher altitudes with those from lower down — and plant them directly in the wild, allowing them to cross-breed naturally and spread traits that might help them survive in the new conditions.

To ensure that seeds are relevant to a climate-changed world, seed banks should team up with climate modellers to predict future vegetation patterns, and focus on keystone species, rare plants or those endemic to particular areas, that are particularly threatened by climate change. Policy-makers must get involved too: a comprehensive international agreement is required to coordinate the collection of genetic material, particularly for cross-border wild species. Finally, seed banks need to think big: rather than collecting 100 grams or so of each variety as is the common practice today, they should be collecting larger amounts, even up to 100 kilograms for some species such as palms with large seeds. This will mean that seed banks of the future will need to be considerably larger to accommodate bulk collections. Only in this way will they capture sufficient genetic diversity to rebuild the world's ecosystems.

Meeting the challenges of seed banking in the twenty-first century will be laborious and expensive, but there is no time to lose if we are to protect the planet's plant life from the rapid environmental change it now faces.

Jeffrey Walck is in the Department of Biology, Middle Tennessee State University, Murfreesboro, Tennessee 37132, USA. Kingsley Dixon is in the School of Plant Biology, the University of Western Australia, and director of science at Kings Park and Botanic Garden, West Perth 6005, Australia.

e-mails: iwalck@mtsu.edu; kdixon@bgpa.wa.gov.au

nature Vol 462|10 December 2009

BOOKS & ARTS

A guide to the day of big data

Michael Nielsen enjoys a rich and stimulating collection of essays on the way in which massive computing power is changing science, from astronomy to zoology.

The Fourth Paradigm: Data-Intensive Scientific Discovery

Edited by Tony Hey, Stewart Tansley and Kristin Tolle

Published at http://research.microsoft.com/en-us/collaboration/fourthparadigm

When it came online in 1946, the US Army's giant ENIAC — Electronic Numerical Integrator and Computer — was hailed as the world's first 'electronic brain', a major step forwards in our ability to process information. It was put to work doing everything from modelling the hydrogen bomb to predicting the weather. Skip to today, and the Large Hadron Collider at CERN, Europe's particle-physics labora-

tory near Geneva, Switzerland, will produce data in a single second that would take, on average, six million ENIACs to store. The Large Synoptic Survey Telescope, planned to begin operation in Chile in 2015, will produce data on a similar scale.

Hundreds of projects in fields ranging from genomics to computational linguistics to astronomy demonstrate a major shift in the scale at which scientific data are taken, and in how they are processed, shared and communicated to the world. Most significantly, there is a shift in how researchers find meaning in data, with sophisticated algorithms and statistical techniques becoming part of the standard scientific toolkit. The Fourth Paradigm is about this shift, how scientists are dealing with it, and some of the consequences. Its 30 chapters, written by some 70 authors, cover a wide range of aspects of data-intensive science.

The book is in four parts. The first two parts are a panorama of the new ways in which data are obtained, through new instruments and large-scale sensor networks. The fields covered range from cosmology to the environment and from healthcare to biology. Most of the chapters in

these sections follow a common pattern. Each introduces a complex system of scientific interest — the human brain, the world's oceans, the global health system and so on — before supplying an explanation of how we are building an instrument or a network of sensors to map out that system comprehensively and, in some cases, to track its real-time behaviour.

We learn in one chapter, for example, about steps towards building a complete map of the human brain — the 'connectome'. Another chapter describes the Ocean Observatories Initiative, a major effort funded by the US National Science Foundation to build an enormous underwater sensor network in the northeast Pacific, off the coasts of Oregon, Washington and British Columbia. And so on, example after example.

enjoyable part of the book. It illuminates common questions that are being asked across these superficially very different fields: who owns the data gathered? How should their release be managed? How should they be curated? How will we preserve them for future generations? Most of all: how can we understand the data?

In parts three and four of the book, these same questions return, from the broader perspective of how the answers could and should

This repetition was, for me, the most

In parts three and four of the book, these same questions return, from the broader perspective of how the answers could and should be reflected in scientific institutions. Part three tackles infrastructure requirements, and part four looks at scholarly communication. Topics include the technical challenges of doing large-scale data analysis, such as multicore and parallel computing; workflow tools that

simplify data analysis and make experiments and analysis more reproducible; and the difficult social and technical challenges of moving to a world in which large data sets are routinely published as part of the scientific process and then integrated with other data sources. The most interesting theme that emerges here is a vision of an increasingly linked web of information: all of the world's scientific knowledge as one big database.

The book has some minor shortcomings. At times, it reads too much like a brochure perhaps inevitable, given that nearly half of the contributing authors come from Microsoft. Many of the essays assume that progress comes mostly from big grants and massive centralized programmes, an assumption not justified by the history of networked innovation. Think of the Internet, or the preprint server arXiv hosted by Cornell University in Ithaca, New York, or the gene-sequence database GenBank — each started by individuals with limited institutional support.

I also found myself wishing that the scope was broader. Science is about more than data: it is about ideas, explanations



A data-storage facility at CERN hints at the huge scale of the information revolution.

and people. The same tools that are driving data-intensive science are also changing the nature of scientific collaboration, and these two changes are closely related. This shift in how scientists team up to create meaning is addressed in only a few chapters.

These are minor criticisms. The rise of 'big data' is one of the major scientific stories of our time, and *The Fourth Paradigm* offers a broad

view that is both informative and stimulating. Better still, the book has been released under a Creative Commons licence, and is available for free on the Internet.

Michael Nielsen is a writer and physicist based in Toronto, Canada. His book *Reinventing Discovery*, about the impact of online tools on science, is due to be published in 2011.

e-mail: mn@michaelnielsen.org

How to get your message across

Don't Be Such A Scientist: Talking Substance in an Age of Style by Randy Olson

Island Press: 2009. 208 pp. \$19.95, £12.99

Am I Making Myself Clear? A Scientist's Guide to Talking to the Public by Cornelia Dean

Harvard University Press: 2009. 288 pp. \$19.95, £14.95, €18.00

The gulf between science and the rest of the world seems to be widening. If you think that keeping your head down, doing your research and not attempting to bridge that gap is enough, two books might convince you that science needs your voice — now.

The first is Randy Olson's *Don't Be Such A Scientist*. Olson was a tenured professor of marine biology at the University of New Hampshire in Durham before packing in his job, packing up his life and moving to Hollywood to learn how to make films. He passes on everything he's learned and saves you the trouble of the embarrassment he experienced as a scientist being cut down to size by film types.

Although the book focuses mainly on making and watching films, it gives some excellent insight into the general areas of communication in which scientists often fail.

Olson has five areas of advice that he uses as chapter headings: 'Don't be so cerebral'; 'Don't be so literal minded';' Don't be such a poor storyteller'; 'Don't be so unlikeable'; 'Be the voice of science!' He advises that scientists need to communicate in broader terms, add some humour once in a while, not shy away from speaking about things in an emotional way, tell interesting stories and be congenial.

Olson gives an excellent explanation of why scientists often have problems communicating with the public, saying that science is a process of "attempting to falsify ideas in the search for truth" and noting that "the masses thrive not on negativity and negation but on positivity and affirmation". He postulates that,

when talking to a general audience, a scientist should try to suppress any urge to be negative because it comes across as arrogant and condescending, something that will often turn an audience against the speaker. This suggestion might not be welcomed by those readers who feel that scientists should never compromise.

Olson believes that science holds the fate of humanity in its hands, and if scientists are incapable of sharing their knowledge with the public then the results could be catastrophic. As more and more people make up their minds about a subject on the basis of a speaker's style, rather than the substance of what they are saying, learning how to speak about science with style is a crucial skill.

The only problem with this book is that the kind of people who need to read it are those who may be most put off by its style. Olson

communicates with his audience in exactly the way he suggests — with humour, emotion and plenty of stories. Some readers may feel that it shouldn't be so much, well, fun.

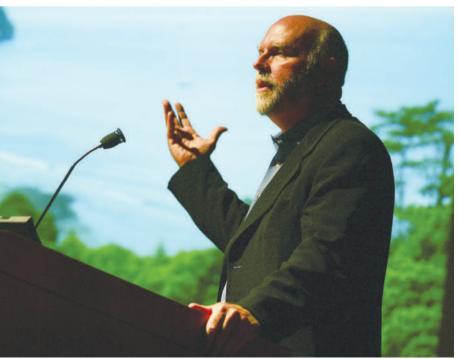
If you want the facts, laid down in a simple, unfussy style, then get a copy of *Am I Making Myself Clear?* by Cornelia Dean, veteran science writer and former science editor of *The New York Times*. This book should sit on the shelf of every scientist, science communicator and university press officer. I've never read a better, more thorough guide to science communication in all its forms.

Dean's suggestions for how to be interviewed by a journalist — for print, radio and television — are spot on. From the preparation you need to do, including how to dress on TV, to always assuming everything you say is 'on the record', her book is packed full of valuable information. She also advises on producing content for the web, writing your own book and press releases, and dealing with politicians.

As Dean puts it: "We need to adopt a broader view of what it means for researchers to fulfill their obligations to society. It is not enough for them to make findings and report them in the scholarly literature. As citizens in a democracy, they must engage, and not just when their funding is at stake."

Gia Milinovich is a science and technology broadcaster for the BBC, Discovery Channel and Channel 4 and a new-media consultant for Hollywood films.

e-mail: giagia@gmail.com



Genomics pioneer J. Craig Venter commands attention through passionate communication.

Third physics opera for Philip Glass

Kepler

Landestheater Linz, Austria

Until 9 January 2010. For performance dates, see go.nature.com/nTDSYn.

As a baritone in the role of astronomer Johannes Kepler sings of his thought processes when discovering that the orbits of planets are ellipses rather than egg-shaped, Philip Glass's music becomes almost lyrical. The renowned minimalist composer uses a rather tuneful set of pieces for the hybrid of science and the arts that is his latest opera, *Kepler*.

The opera was commissioned as part of the Linz 2009 European Capital of Culture celebrations by the Landestheater in Linz, Austria, the city where Kepler laboured for many years after moving from Prague in 1612, where he had succeeded Tycho Brahe as court mathematician. We saw the concert version, shown some two months after the opera's Austrian premiere, as part of the Next Wave Festival at New York's Brooklyn Academy of Music.

As Glass explained to us in a personal interview, the opera's commissioning gave him the Austrian librettist Martina Winkel. To provide historical context for Kepler's stay in Linz, she combined phrases from Kepler's extensive writings with jarring lines of poetry written by the seventeenth-century German poet Andreas Gryphius. These chart the death and destruction of the Thirty Years War, which was being waged in Germanic lands at the time.

This is Glass's third opera about a physicist, following on from Einstein on the Beach (1976) and Galileo Galilei (2002). Here the composer decided not to present Kepler's life but to focus instead on his ideas. As he noted in a public discussion with US physicist Michio Kaku before one performance, Kepler's thoughts advanced from his early geometrical ideas to the less easily visualized calculations that brought his success. That said, the libretto tosses around the term 'equation' perhaps too glibly given what Kepler actually did.

The opera begins with a brief, rousing overture and runs for 115 minutes without an intermission. Kepler's epitaph, "Once I measured the heavens/Now I measure earthly shadows," begins and ends the work. In the concert version, four basses and a dozen cellos provide the only warmth of colour against an otherwise black stage. The energetic music thrusts the action forward as the 38-member chorus, six anonymous soloists and Kepler — sung by Martin Achrainer — perform their parts. Glass has resisted the temptation to



Baritone Martin Achrainer in the role of Johannes Kepler, on stage in Linz, Austria.

"The renowned minimalist

composer uses a rather tuneful set

of pieces for the hybrid of science

and the arts that is his latest opera."

assign the six anonymous soloists to each of the six planets known in Kepler's time, or to try to translate the planetary orbital periods directly into the musical notes or harmonies that they imply. Instead, swelling arpeggios in the extensive string section, together with varied use of percussion from a maraca to gongs

and drums, make for a stimulating and engrossing work.

Kepler's groundbreaking theory of elliptical orbits from 1605 (published in

1609 in his book *Astronomia Nova*; see page 725) appears only in the last half-hour of the performance. The harmonic law that he found ten years later is not included. Neither are the specific planetary calculations of his Rudolphine Tables or their verification through the observation of a transit of Mercury. We wished that astronomers or historians of astronomy had been consulted at an early stage of writing and been given an opportunity to make suggestions.

Perhaps Glass's opera will succeed in introducing Kepler and his theories on the Solar System to a wider public. Unlike Galileo his contemporary, Kepler is not well known among non-scientists, so the more information about him that seeps into general culture, the better. After all, who would know today about the malfeasance of Simon Boccanegra as the fourteenth-century Doge of Genoa were it not for his immortalization in Verdi's opera?

It is a shame that celebrations for the International Year of Astronomy have not included

a revival of Glass's *Galileo Galilei* on any major stage, even though 2009 marks the 400th anniversary of Galileo's first recorded use of a

telescope to survey the heavens. Neither has Paul Hindemith's 1957 opera about Kepler, *Die Harmonie der Welt*, had an airing. Kepler fans such as the International Astronomical Union's new working group about Kepler, who feel that the proper 400th anniversary to note this year is that of the publication of *Astronomia Nova*, will be pleased that the occasion is honoured with this opera in his name.

Jay M. Pasachoff is director of the Hopkins Observatory, Williams College. Naomi Pasachoff is a research associate at Williams College, Williamstown, Massachusetts O1267, USA. e-mails: jay.m.pasachoff@williams.edu; naomi.pasachoff@williams.edu

In Retrospect: Kepler's Astronomia Nova

Jack J. Lissauer explains how the great astronomer's insight into planetary orbits is still revealing new views of the Universe four centuries on — from extrasolar Earths to black holes.

Astronomia Nova by Johannes Kepler First published 1609.

The 2009 International Year of Astronomy commemorates the 400th anniversary of two great discoveries: Galileo Galilei's detection of Jupiter's four major moons and the publication of Johannes Kepler's monumental book Astronomia Nova. In this 1609 work, Kepler demonstrated that planets follow elliptical paths about the Sun, a breakthrough that set up his famous three laws of planetary motion, still fundamental to physics and astronomy today.

Astronomia Nova introduced the first two of Kepler's three laws: first, that each planet moves on an elliptical path, with the Sun at one focus; and second, that a line connecting any planet to the Sun sweeps out area at a constant rate — which implies that a planet moves faster when it is nearer the Sun. In 1618, he deduced his third law, which relates the motions of the planets to one another: the ratio of the orbital period squared and the mean distance from the Sun cubed is the same for each planet.

The full title of Kepler's treatise, in translation, notes the significant role that measurements of the orbit of Mars had in the deduction of his laws: New Astronomy Based upon Causes, or Celestial Physics Treated by Means of Commentaries on the Motions of the Star Mars, from the Observations of Tycho Brahe, Gent. Observations of the red planet were key because its orbit is more eccentric than any other planet apart from Mercury, which is close to the Sun and difficult to observe. Kepler was also familiar with Mars because he was given accurate data on its motion when he was hired to assist the great astronomer Tycho Brahe in 1600. The inclusion of 'physics' in the title signifies Kepler's mathematically and physically rigorous approach to the problem one that is familiar to any modern scientist but not to natural philosophers of his era. Kepler's thinking was far ahead of that of most of his contemporaries.

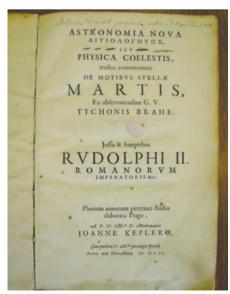
In the same year that *Astronomia Nova* was published, the heliocentric model of the Solar System was only just verified. The concept that Earth and other planets orbit the Sun is credited primarily to Nicolaus Copernicus, who published his theory in 1543, more than 60 years before Kepler's book, although the idea itself dates back at least as far as the Greek

astronomer and mathematician Aristarchus of Samos, who lived in the third century BC. Galileo's discovery of four moons orbiting Jupiter using a telescope confirmed that Earth wasn't the centre of the Universe around which everything in the celestial domain revolves.

Kepler was the first to realize that orbits are better described by ellipses traversed at nonuniform speed than by combinations of 'perfect' circular motions on which orbits had hitherto been modelled. Like a modern data analyst, he incorporated all of the observations with their inherent measurement errors in an attempt to fit a model. His weaving together of observations and mathematical models into a physical theory was a vanguard for future studies.

Kepler's revolutionary model was accurate enough to predict the transits of the inner planets Mercury and Venus across the Sun's disc, which were first observed in 1631 and 1639, respectively. Thus it is fitting that Kepler's name today graces the first space mission dedicated to searching for planets beyond our Solar System using a similar method. Launched on 7 March this year, the Kepler spacecraft's telescope will continuously monitor the light from more than 100,000 stars looking for planetary transits over a period of at least three-and-a-half years. When a planet passes between its star and the telescope, once per orbit, the star dims measurably.

Because large planets orbiting close to their



Johannes Kepler's 1609 book outlined his first two laws of planetary motion.

stars block more light, such planets are easiest to detect using the transit method. Hence, to date, all of the planets discovered in this way from the ground are larger than Uranus and have orbital periods of less than 10 days. But the clear view and unbroken observations available from space mean that the Kepler mission should detect smaller and more distant planets, notably Earthlike bodies with one-year orbital periods around Sun-like stars. Even smaller planets orbiting near their stars will be detectable, as will more distant planets that are larger than Earth.

Since Kepler deduced his laws, there have been many advances in our knowledge of planetary orbits, Isaac Newton's theories of motion and gravity being the most substantial. The trajectories of planets — and the Sun — can be perturbed by the gravitational pulls of other planets and moons. Deviations of the motion of Uranus from that explained by Newton's model led to the discovery of Neptune in 1846. By contrast, irregularities in the motion of Mercury were explained only by a completely new theory of space-time and gravity, Albert Einstein's general theory of relativity, published in 1915.

As astronomy has expanded its horizons, the application of Kepler's laws has led to even stranger discoveries. Invisible dark matter is needed to explain the fast orbits of gas in the outskirts of spiral galaxies - these orbital speeds exceed the predictions of Kepler's third law, which implies that an object with a large orbit should move slowly. And, by contrast, the Keplerian behaviour of the orbits of stars $\frac{2}{5}$ in our Galaxy's centre has revealed the existence of a supermassive black hole at the core of the Milky Way.

Kepler was an early champion of the application of reason to our understanding of the Universe. He wrote in his introduction to Astronomia Nova: "I prove philosophically not only that the Earth is round ... not only that it is contemptibly small, but also that it is carried along among the stars." Four hundred years later, it is striking that we are still using his laws to measure planetary systems that lie far from

Jack J. Lissauer is a space scientist at NASA's Ames Research Center, Space Science & Astrobiology Division, Moffett Field, California 94035, USA, and a visiting fellow at the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK.

e-mail: Jack.J.Lissauer@nasa.gov

DEVELOPMENTAL BIOLOGY

Asymmetry with a twist

Nipam H. Patel

In snails, manipulating the orientation of cells in the early embryo alters the left-right asymmetry of the shell and body. These findings refine the search for the symmetry-breaking event in this and other animals.

The elegant shape of seashells has long fascinated us, and their beautiful curved forms have inspired many man-made designs. The spiral shapes can be divided into two distinct groups on the basis of the direction of coiling. Hold a shell so that you look down onto the pointed end, and you will see that most shells spiral downward in a clockwise direction (Fig. 1); only a few twist in the opposite direction. This property of shell coiling is termed chirality; shells with clockwise spirals are called dextral shells, and those with anticlockwise spirals are termed sinistral shells.

These differences in chirality are a manifestation of left-right asymmetry of the organs of the animal. In the case of the shell, the gland that makes this structure is producing shell material at different rates on the left and right sides, leading to the coiling pattern as the shell grows. Through a conceptually simple, although technically challenging, study in the snail, Kuroda and colleagues1 show (page 790 of this issue) that they can reverse the chirality of the shell, and the asymmetry of other organs, by mechanically altering the relative orientation of cells in the early embryo.

Kuroda *et al.* worked with a pond snail, *Lym*naea stagnalis. Whereas most species of snail have dextral shells, and a few have sinistral shells, L. stagnalis falls into a relatively rare group in which both chiralities exist within the same population. The difference between dextral and sinistral individuals can be detected during the first few cell divisions of the embryo — as the four-cell embryo undergoes cleavage to become an eight-cell embryo. For dextral snail embryos, the smaller daughter cells at the four-to-eightcell transition twist clockwise as cell division occurs. In the case of sinistral snail embryos, the daughter cells are initially directly on top of their division partners, but then twist anticlockwise (Fig. 2, overleaf). The result is that dextral and sinistral embryos look like mirror images of each other at the eight-cell stage, just as their shells will look like mirror images of each other as adults.

Thanks to the outstanding work of 'gentleman' scientists of the 1920s and 1930s, we also understand the genetics of snail chirality.



Figure 1 | Spiralling downward. Most shells, like that of the sea snail Architectonica perspectiva shown here, spiral downward in a clockwise manner and are termed dextral shells. Shells with anticlockwise spirals are called sinistral shells.

Professor Arthur Boycott of the University of London recruited a team of naturalists, most notably Captain Cyril Diver, a part-time parliamentary administrator, to examine the genetics of chirality in *Lymnaea peregra*, a close relative of L. stagnalis. What they described² was the founding example of the phenomenon we now call maternal-effect inheritance³. Simply put, it is not the genotype of the individual animal that determines a particular phenotype (in this case the dextral or sinistral shell); rather, it is the genotype of the mother that determines the phenotype of her progeny. In the case of both Lymnaea species, a single genetic locus controls chirality, with the dextral allele acting in a dominant manner. If we think of the two alleles of the chirality locus as D and d, then mothers who are DD or Dd give rise to dextral progeny, whereas dd mothers give rise to sinistral progeny. The genotype of the father has no bearing on the phenotype of the offspring.

Thus, Kuroda et al. started with DD or *Dd* mothers, knowing that all their progeny would exhibit a dextral cleavage pattern and normally form dextral shells, and with dd mothers, whose embryos would display a sinistral cleavage pattern and form sinistral shells. By using small glass rods to push on the cells, the authors could change the cleavage pat-

tern of the embryos. They took embryos that

should have cleaved dextrally, and forced

the cells at the eight-cell stage to take on

the orientation normally seen for sinistral embryos, and vice versa (Fig. 2). Remarkably, this reversed the chirality of the animals as they grew into adults. So, merely changing the relative orientation of the cells at the eight-cell stage can completely override the maternal-effect specification of chirality. Before these experiments¹, it was conceivable that the chirality of the cell-cleavage pattern and that of the shell and organs later on were correlated, but did not represent cause and effect. Now we know that the maternal chirality locus controls the orientation of the cells at the eight-cell stage, and that this orientation ultimately controls the adult snail's chirality.

Kuroda and colleagues' findings¹ suggest that it is the pattern of cell-cell interactions initiated at the eight-cell stage that dictates shell chirality, and that these interactions differ depending on the cleavage pattern. We still do not know what the maternal chirality locus encodes, but the gene product in some way regulates the cellular cytoskeleton and thus dictates the orientation of cell division and of cells at the eight-cell stage. We also do not know the nature of the subsequent cellular interactions, but we might glean some clues from studies of the nematode worm Caenorhabditis elegans. Similar mechanical manipulations were used to show that specific cell-cell interactions establish left-right asymmetry in C. elegans⁴, and genetic screens⁵ have now revealed some of the molecules involved in this process.

Of course, this type of left-right asymmetry is not confined to snails and nematodes. For example, humans and other vertebrates have striking asymmetries in the placement of organs such as the heart, lungs, liver and gut. Recent work⁶ has shown that the genes used to control left-right asymmetry in snails and vertebrates share certain features. Specifically, the genes encoding the signalling molecule Nodal

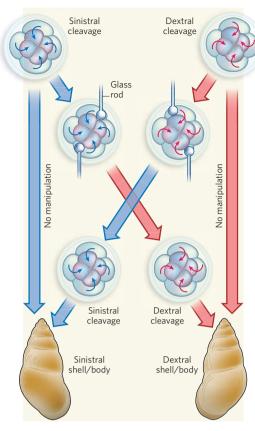


Figure 2 | Reversing asymmetry in snails. At the transition from the four- to eight-cell stage in the embryo of the pond snail, Lymnaea stagnalis, newly formed smaller daughter cells twist either clockwise or anticlockwise. This chirality (direction of twisting) correlates with the chirality of the adult shell and the left-right asymmetry of other body organs. For sinistral snail embryos, the smaller daughter cells twist anticlockwise after cell division (sinistral cleavage; blue arrows). For dextral snail embryos, the daughter cells twist clockwise as cell division occurs (dextral cleavage; red arrows). Kuroda et al. used glass rods to push on the cells at the eight-cell stage, forcing them into the opposite orientation. The result was a reversal of chirality of the individual that developed from the manipulated embryo. Embryos that started with a sinistral cleavage, but were forced into a dextral orientation, grew into adults with dextral shells, whereas embryos that started with a dextral cleavage pattern, but were forced into a sinistral orientation, developed into adults with sinistral shells.

and the transcription factor Pitx, both of which have well-studied roles in vertebrate left-right asymmetry, are asymmetrically expressed (only on the right side in dextral species and only on the left side in sinistral species), and are functionally involved in determining left-right asymmetry of snails starting at about the 32–64-cell stage⁶. Indeed, the mechanical manipulations of Kuroda *et al.*¹ also reversed the left-right expression of these two genes.

Thus, studying snail left-right asymmetry will be relevant to understanding the phenomenon in vertebrates as well. In mice, the

symmetry-breaking event seems to involve an asymmetry in extraembryonic-fluid flow set up by cilia⁷, whereas in chickens, it seems instead to involve early embryonic cell migration⁸. Snail embryos do not have cilia at the eight-cell stage, so it is difficult

to predict how the maternally controlled symmetry-breaking process in snails relates to the early events in vertebrates. The final twist to this tale is yet to be told.

Nipam H. Patel is in the Department of Molecular Cell Biology and the Department of Integrative

Biology, University of California, Berkeley, 519A LSA 3200, Berkeley, California 94720-3200, USA. e-mail: nipam@uclink.berkeley.edu

- Kuroda, R., Endo, B., Masanori, A. & Shimuzu, M. Nature 462, 790-794 (2009).
- 2. Boycott, A. E. & Diver, C. Nature 119, 9 (1927).
- 3. Gurdon, J. B. Cell 123, 751-753 (2005).
- 4. Wood, W. B. Nature 349, 536-538 (1991).
- 5. Sarin, S. et al. Genetics 176, 2109-2130 (2007)
- Grande, C. & Patel, N. H. *Nature* **457**, 1007–1011 (2009).
 Hirokawa, N., Tanaka, Y., Okada, Y. & Takeda, S. *Cell* **125**.
- Hirokawa, N., Tanaka, Y., Okada, Y. & Takeda, S. Cell 125, 33-45 (2006).
- 8. Gros, J. et al. Science 324, 941-944 (2009).

GAMMA-RAY BURSTS

Magnetism in a cosmic blast

Maxim Lyutikov

Astronomers know little about γ -ray bursts other than that they are the most energetic explosions in the Universe. The latest observations indicate that large-scale magnetism contributes to their power.

Gamma-ray bursts (GRBs) are truly amazing astrophysical events. In a matter of seconds, they release more energy than the Sun will do in its lifetime of 10 billion years. These powerful stellar explosions eject material at highly relativistic velocities — differing from the velocity of light by as little as one part in a million. Because they are so bright, GRBs can be seen almost to the edge of the observable Universe, occurring, on average, about twice a day. However, their origin remains a mystery. Their outflows could be gas-dynamic phenomena, driven by gases near their central engine with a high kinetic pressure (to some extent, similar to exhaust jets in aeroplanes)1. Alternatively, and somewhat unexpectedly, magnetic fields, which on Earth have a negligible dynamic role, have been proposed as the dominant mechanism for both driving the bulk of the outflow and accelerating particles within it^{2,3}. The coupling of magnetic fields to gravity — which occurs, for example, in the close vicinity of the event horizon (a boundary beyond which nothing can escape) of a rotating black hole through the Blandford-Znajek mechanism⁴ — can indeed produce powerful relativistic outflows such as those seen in GRBs.

But magnetic fields in GRBs are notoriously hard to observe — just imagine the difficulties faced by an observer trying to prove the existence of Earth's magnetic field without the use of a magnetic compass. On page 767 of this issue, Steele *et al.*⁵ report a possible detection of magnetic fields in a GRB, GRB 090102, through observations of polarization in its optical (visible-light) emission.

GRBs emit light across a broad band of frequencies, ranging from the radio to the high-energy end of the γ -ray part of the electromagnetic spectrum. As the name suggests, they are detected, at least initially, through their

 $\gamma\text{-ray}$ radiation. Because Earth's atmosphere effectively blocks $\gamma\text{-rays}$, thereby protecting the planet from this highly damaging radiation, astronomers rely on space telescopes that operate at high frequencies, such as Swift, to observe them. Once a GRB is detected by a space telescope, its sky coordinates are transmitted to ground-based observatories, which then carry out follow-up observations.

In the optical waveband, GRBs tend to produce a dim flash that lasts for only a few dozen seconds. This makes ground-based follow-up observations challenging. First, the telescope must turn automatically towards a GRB, because any delay may mean that the quickly fading signal is not detected. Second, because the telescope sometimes needs to be reoriented by a large angle during this 'slewing' manoeuvre, modern, large telescopes are often too massive, and therefore too slow, to be useful.

In their study, Steele and colleagues⁵ used the medium (2-metre) RINGO Liverpool Telescope at La Palma in the Canary Islands. The key advantage of the RINGO telescope is that it is equipped with a detector that can measure the polarization of incoming radiation. The authors were therefore able to measure a considerable degree of polarization — of the order of 10% — in the optical emission from GRB 090102. The fraction of polarization in astrophysical sources rarely exceeds several per cent.

Polarization is a property of electromagnetic waves that describes the preferred direction of their electric-field oscillations. A non-zero polarization indicates that the process that produced the waves is, in some sense, non-isotropic: it is more sensitive to one particular direction than others. The most logical — but not the only — explanation for the high degree of polarization obtained by Steele and

colleagues is that the emitting source of GRB 090102 is permeated by large-scale, ordered magnetic fields, and that the emission process is non-thermal synchrotron emission from leptons (such as electrons) as they spiral around the magnetic fields at high (relativistic) speeds in the GRB outflow's frame of reference.

Magnetic fields are necessary to produce synchrotron emission. But if a magnetic field is tangled on small scales, as the favoured model of GRB emission posits, the polarization produced by different emitting regions will average out to nearly zero. A large amount of polarization would therefore be tied to large-scale, non-tangled magnetic fields. If this interpretation is correct, Steele and colleagues' result will lend support to models in which large-scale magnetic fields have an important role in launching and collimating GRB outflows. In addition, large-scale magnetic fields can directly accelerate the emission of particles through the process of magnetic reconnection — a well-established hypothesis to explain solar flares.

Steele and colleagues' detection⁵ of polarized light from GRB 090102 is likely to contribute to the heated debate about the nature of GRBs. Previously, claims⁶ of high polarization

in GRBs were inconclusive⁷ because polarization is difficult to measure at high frequencies, giving large uncertainties. At visible-light frequencies, polarization can be measured with much higher certainty. A conservative interpretation of the results of Steele *et al.* is that the magnetic forces in GRBs are at least as important as the gas-pressure forces. A more exciting possibility is that magnetic fields completely dominate the outflow dynamics, so that the nature of GRBs is mostly electromagnetic and not gas-dynamic⁸.

Maxim Lyutikov is in the Department of Physics, Purdue University, West Lafayette, Indiana 47907-2036. USA.

e-mail: lyutikov@purdue.edu

- 1. Piran, T. Rev. Mod. Phys. 76, 1143-1210 (2004).
- 2. Lyutikov, M. New J. Phys. **8**, 119 (2006).
- 3. Usov, V. V. Nature 357, 472-474 (1992).
- Blandford, R. D. & Znajek, R. L. Mon. Not. R. Astron. Soc. 179, 433-456 (1977).
- Steele, I. A., Mundell, C. G., Smith, R. J., Kobayashi, S. & Guidorzi, C. Nature 462, 767-769 (2009).
- 6. Coburn, W. & Boggs, S. E. Nature 423, 415-417 (2003).
- Rutledge, R. E. & Fox, D. B. Mon. Not. R. Astron. Soc. 350, 1288–1300 (2004).
- Blandford, R. D. in Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology (eds Gilfanov, M., Sunyeav, R. & Churazov, E.) 381-404 (Springer, 2002).



Excitatory view of a receptor

Lonnie P. Wollmuth and Stephen F. Traynelis

Ion channels opened by glutamate mediate fast cell-to-cell information transfer in the nervous system. The structure of a full-length tetrameric glutamate receptor is both confirmatory and revelatory.

Anyone who wears glasses knows the experience of putting them on. What was fuzzy, blurry and ill-defined suddenly snaps into focus. On page 745 of this issue, Sobolevsky *et al.*¹ offer those working on signalling in the nervous system, and on membrane proteins in general, a pair of glasses. The authors provide the first structure of a full-length glutamate receptor, giving a spectacular, panoramic view of a surprising landscape with at least a few unanticipated features.

Fast cell-to-cell signalling in the nervous system — the basis of our ability to perceive, think and respond — occurs at specialized structures called synapses, sites where encoded information is transferred from one cell to another by a chemical neurotransmitter. In the nervous system, glutamate is the major excitatory neurotransmitter, binding to and activating ionotropic glutamate receptors. These receptors are transmembrane proteins that have a glutamate-recognition site (ligand-binding domain) that, when bound by glutamate, opens an associated ion channel. They are therefore at the heart of nervous-system function. Regrettably,

when unregulated, they can also contribute to an array of debilitating disorders, including schizophrenia, Alzheimer's disease and Parkinson's disease, and are involved in the neuronal damage that accompanies stroke and traumatic brain injury.

The structure presented by Sobolevsky et al. is of one of the main glutamate-receptor subtypes, an AMPA (α-amino-3-hydroxy-5methyl-4-isoxazole propionic acid) receptor from the rat. It is made up of four GluA2 subunits (GluR2 in older nomenclature) that are identical in terms of amino-acid sequence, and it encompasses three structural/functional domains (Fig. 1a, overleaf). Two of these domains are located on the external side of the cell membrane, and we have seen them individually before — the modulatory amino-terminal domain (ATD)²⁻⁴ and the ligand-binding domain (LBD)5 with its clamshell-like arrangement. The third component is the transmembrane domain (TMD), which forms the ion channel, and this is our first view of it. In many ways the structure is comforting, because it consolidates and verifies much



50 YEARS AGO

Wisdom of the West. By Bertrand Russell — This is Lord Russell's second brilliant venture into a comprehensive history of Western (mainly philosophical) thought ... Of course, despite the brilliance and apparent novelty of presentation, it should be realized that on the whole this is old wine poured into new bottles ... There are still the rough and ready divisions into the Continental rationalists (who are rather wicked) and the more virtuous British empiricists, supported arbitrarily by the quite unproved suggestion that the diagrams representing these two types of philosophical systems are, respectively, a pyramid standing on its head, as against the other, standing on its feet ... But these are minor criticisms. Certainly this new excursion into our cultural heritage is another astonishing venture testifying to the brilliance and almost legendary energy of its author. From Nature 12 December 1959.

100 YEARS AGO

In his review of Prof. Poulton's work, "Charles Darwin and the Origin of the Species," Prof. Meldola says ... that the Darwinian theory is absolutely dependent upon the truth of the belief "in the transmissibility by inheritance of individual differences or 'fluctuations'." ... There is now available a vast amount of evidence tending to show that "fluctuations" seemingly the direct results of changes in the environment are inherited; but how is it possible to convince Weismann and his followers that such "fluctuations" have not been due, as they will say, to "spontaneous germinal variations"? Surely the onus probandi really rests with them! We have here the question of the inheritance or not of acquired characters reduced to its simplest terms.... What evidence can those who disbelieve in the inheritance of acquired characters present to show that in all such cases there *must* be a primary germinal

From Nature 9 December 1909.

& 100 YEARS AG

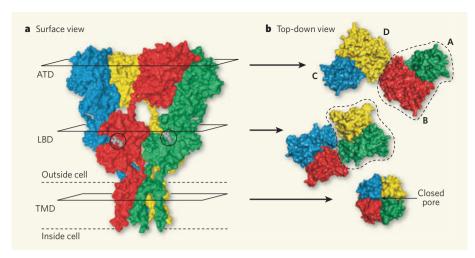


Figure 1 | Structure of a full-length tetrameric glutamate receptor. Sobolevsky et al.¹ report the first X-ray structure of a full-length glutamate receptor (minus the carboxy-terminal domain), specifically of the AMPA receptor subtype. a, Surface representation of the AMPA receptor with the four subunits of identical amino-acid sequence (A–D) coloured green (A), red (B), blue (C) and yellow (D). The amino-terminal domain (ATD) and ligand-binding domain (LBD) residing on the external side of the cell membrane, and the transmembrane domain (TMD) that forms the ion channel, are indicated. The circles indicate competitive antagonists (grey) occupying the agonist-recognition sites. b, Top-down view of the ATD, LBD and TMD, illustrating the domain swapping and symmetry mismatch (between LBD and TMD). The dashed line for the ATD and LBD indicates the dimer containing the A subunit, which in the ATD associates with the B subunit, but in the LBD associates with the D subunit. The TMD shows four-fold symmetry. Because the LBD is bound by an antagonist, the permeation pore, located in the centre of the TMD, is closed.

previous functional and structural work. But at the same time it is exhilarating, owing to the unexpected way in which these domains are intertwined and linked together.

The ATD and LBD in the new structure overlap remarkably well with earlier, isolated structures $^{2-5}$ of these domains. They are arranged as dimers, a key structural/functional motif for receptor function; the agonist (glutamate) recognition site, in this instance occupied by a competitive antagonist, is located within the clam shell formed by the LBD. These domains also have a two-fold symmetry relative to the axis perpendicular to the cell membrane. Conversely, the TMD has four-fold symmetry, which is perhaps not surprising given its kinship to another kind of transmembrane ion channel, the potassium (K^+) channel.

A completely unanticipated feature of the tetrameric GluA2 receptor is that domain swapping and crossover occurs between subunits. As a result, the homotetrameric GluA2 protein complex has two conformationally distinct pairs of subunits, referred to as A/C and B/D (Fig. 1b). Thus, at the level of the ATD, the dimer pairs are A–B and C–D, with considerable inter-pair interactions between the B and D subunits; at the level of the LBD, however, the dimer pairs are A-D and B-C, with inter-pair interactions occurring between A and C (Fig. 1b). This pairwise arrangement is abolished in the TMD, in which four independent but equivalent subunits have four-fold symmetry.

Sobolevsky and colleagues' work¹ also gives us our first glimpse of the glutamate receptor's transmembrane ion channel. Functional

evidence has supported the once radical idea^{6,7} that the core of the glutamate-receptor ion channel — transmembrane helix M1, the M2 pore loop and transmembrane helix M3 — shares structural similarity and perhaps evolutionary homology with the permeation pore in K⁺ channels⁸, an idea overwhelmingly supported by the new structure. Moreover, glutamate receptors also reprise another feature of K⁺ channels: they have an additional peripheral transmembrane helix, the M4 segment, that associates with the ion-channel core of an adjacent subunit, as do peripheral transmembrane helices in voltage-gated K⁺ channels. The significance of this arrangement in glutamate receptors is unknown. It may, however, represent a common structural theme in transmembrane proteins and adds a further intriguing aspect to the evolutionary history of glutamate receptors.

In the new structure¹, the LBD is occupied by a competitive antagonist, so the status of the associated ion channel is not controversial — it is in a closed, non-conducting state. As predicted from previous experimental results^{9,10}, highly conserved residues at the apex of the transmembrane gating helix M3 (like the inner helices in a K⁺ channel) are positioned in close proximity and presumably form the gate that blocks ion permeation through the closed channel. A clear definition of the channel gate is an essential step forward in defining receptor function, and will stimulate experiments that further focus our view of channel activation.

Earlier images of glutamate-receptor fragments/domains were missing the two sets of linkers that couple the ATD to the LBD and the LBD to the TMD. Yet it is these linkers that accommodate both the domain swapping and symmetry mismatch, and ultimately transduce ATD modulatory effects to the LBD/TMD and the conformational change in the LBD to channel opening and closing. Indeed, visualization of the details reveals previously unknown linker arrangements — especially for those coupling the LBD to the ion channel, which, as Sobolevsky et al. propose, may be essential for receptor function. For example, the linker preceding the M1 transmembrane helix (the pre-M1 region) makes a short helix parallel to the plane of the membrane and contacts the protruding carboxy- and amino-terminal ends of transmembrane helices M3 and M4, respectively. Closure of the LBD clamshell presumably moves this pre-M1 helix away from the M3 helix, permitting this main transmembrane gating element to rotate away from the central axis of the pore and opening the ion channel. Thus, the new structure hints at potential gating steps¹¹ that may be functionally distinct. It allows specific predictions to be made that could transform kinetic studies of glutamate receptors from an impersonal mathematical exercise to one that also involves matchmaking between elementary kinetic steps and protein conformations.

The different conformations for identical subunits within a tetrameric protein were unanticipated. However, another glutamate receptor subtype, the NMDA (N-methyl-Daspartate) receptor, requires two different types of subunit to function — typically two GluN1 and two GluN2 subunits, although their exact arrangement was unknown. Using a biochemical approach, and taking advantage of the two conformationally distinct sets of subunits, Sobolevsky et al. provide evidence that the GluN1 subunits take the A-C role and the GluN2 subunits take the B-D role, and that the subunit arrangement is GluN1-GluN2-GluN1-GluN2. This is a provocative result and seems rational given the need for different subunits.

It is easy to be excited by these data, and Sobolevsky et al. should indeed be commended for their groundbreaking work, which will shift studies of glutamate receptors in a fresh direction. But, of course, more work is needed. A deeper understanding of receptor function will require not only clever functional experiments inspired and guided by structure, but also structures of higher resolution, snapshots of alternative states, and structures of other members of the glutamatereceptor family. We should also retain a healthy dose of caution, lest we rush to transfer all the principles learned here prematurely to the entire glutamate-receptor family. However, for the moment, enjoy the newly afforded clearer view.

Lonnie P. Wollmuth is in the Department of Neurobiology and Behavior, and the Center for Nervous System Disorders, Stony Brook University, Stony Brook, New York 11794-5230, USA. Stephen F. Traynelis is in the Department of Pharmacology, Emory University School of Medicine, Rollins Research Center, Atlanta, Georgia 30322-3090, USA. e-mails: lwollmuth@ms.cc.sunysb.edu; strayne@emory.edu

- Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. Nature 462, 745–756 (2009)
- 2. Jin, R. et al. EMBO J. 28, 1812-1823 (2009).
- 3. Karakas, E., Simorowski, N. & Furukawa, H. *EMBO J.* doi:10.1038/emboj.2009.338 (2009).
- 4. Kumar, J., Schuck, P., Jin, R. & Mayer, M. L. Nature Struct.

- Mol. Biol. 16, 631-638 (2009).
- Armstrong, N., Sun, Y., Chen, G.-Q. & Gouaux, E. Nature 395, 913–917 (1998).
- Wo, Z. G. & Oswald, R. E. Trends Neurosci. 18, 161–168 (1995).
- 7. Wood, M. W., VanDongen, H. M. & VanDongen, A. M. *Proc. Natl Acad. Sci. USA* **92**, 4882–4886 (1995).
- Chen, G.-Q., Cui, C., Mayer, M. L. & Gouaux, E. Nature 402, 817–821 (1999).
- Qian, A. & Johnson, J. W. Physiol. Behav. 77, 577-582 (2002).
 Chang, H.-R. & Kuo, C.-C. J. Neurosci. 28, 1546-1556
- 11. Banke, T. G. & Traynelis, S. F. *Nature Neurosci.* **6,** 144–152

boundary, making the results calibration-independent. The upshot is that the authors conclude that the two phases, perovskite and post-perovskite, could coexist over a depth range of $400-600~(\pm 100)~km$ — a much larger range than the seismically estimated D" discontinuity thickness of about 30 km (ref. 7).

How can this discrepancy between a thin D" discontinuity (that is, a seismically sharp boundary) and the results of Catalli et al. be explained? The authors contend that compositional differences, perhaps low aluminium content or high abundance of ferropericlase (Mg, Fe)O — Earth's second most abundant mineral — would sharpen the boundary. But this would run counter to the view that portions of the D" layer contain some oceanic crust, subducted from the surface through plate tectonics. Although oceanic crust mineralogy would have a large proportion of magnesium silicate perovskite, it also has increased amounts of silica (SiO₂) and alumina (Al₂O₃), which, given Catalli and colleagues' results, suggests an even thicker D" discontinuity.

All in all, the D" layer presents many puzzles. One, for instance, is that seismic data indicate that there must be directionality to the minerals present, possibly due to the planar crystal structure of post-perovskite. But there is no consensus as to how post-perovskite would provide a consistent mechanism for such directionality^{8,9}. Another puzzle concerns a region at the very bottom of D" that is dubbed the ultra-low-velocity zone, where large decreases in the speeds of seismic waves are recorded. One explanation for this phenomenon could be the presence of a small amount of minerals in melted form; another explanation is the possible existence of iron-rich post-perovskite produced by reaction with the iron-rich core¹⁰. Indeed, the D" region is evidently chemically diverse — whether owing to interactions with the core, old oceanic crust or other factors and this is likely to be the root cause of much of its unusual seismic behaviour.

There is much left to pin down in understanding the causes of the seismic signatures in D". Catalli and colleagues' results highlight the need for more experiments and computations on the behaviour of not just Earth-relevant compositions of magnesium silicate perovskite, but also of other mineral assemblages, to tease out the effects of chemical composition on the transition to post-perovskite. Quantifying the effects of melt in D" on density and on seismic wave speeds will also be important to understanding the other puzzles presented by D". In addition to new lab results and sophisticated computations, fine tuning on the D" features that can be resolved by seismological observations are also needed.

Kanani K. M. Lee is in the Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06511, USA. e-mail: kanani.lee@yale.edu

EARTH SCIENCE

The enigma of D"

Kanani K. M. Lee

A phase transition of Earth's most abundant mineral occurs at pressures and temperatures corresponding to those thought to exist just above Earth's core. New experiments shed light on this enigmatic D" region.

At the half-way point in a journey to Earth's centre, at a depth of about 2,900 kilometres, an intrepid explorer would encounter the coremantle boundary. This is where Earth's rocky mantle meets the fluid, iron-rich outer core, and it is marked by a large change in density and chemistry. The boundary is also thermal, because Earth's core is more than 1,000 K hotter than the mantle¹. Thus, conduction is the characteristic form of heat transfer for some 100 kilometres or so immediately above the core, whereas above that, in the bulk of the mantle, convection dominates.

The sharp change in temperature gradient creates a distinct seismological and mineralogical environment for this region, which is dubbed D" (D double prime). There is, of course, no explorer to send back reports about the region, so the study of D" depends on the analysis of seismic waves or on experiments on minerals at relevant temperatures and pressures. The latter is the approach taken by Catalli $et\ al.\ (page\ 782\ of\ this\ issue)^2,$ who have come up with a new estimate for the thickness of the D" discontinuity and the nature of this region.

The special mineralogy of D" is produced by a transition in crystal structure from that of magnesium silicate perovskite, MgSiO₃, the most abundant mineral on Earth, to a form known as post-perovskite^{3,4}. Seismicwave speeds change abruptly in this region, often more than once⁵, providing information for seismological investigation of D". The approach taken by Catalli *et al.*² instead involves experimental determination of a feature called the Clapeyron slope, which marks the coexistence curve between two mineral phases, and is defined by the change in temperature over the change in pressure. For the perovskite-to-post-perovskite transition, this

slope is positive — that is, as one goes deeper into the mantle to higher pressures, higher temperatures are required to produce postperovskite. In addition, for a material that varies in composition (for example, with iron and aluminium replacing some of the magnesium), there is a finite range of pressures and temperatures for which both the high-pressure and low-pressure phases coexist; thus a 'thickness' of this phase transition will yield a boundary that is either sharp or broad.

Catalli and co-authors have made precise measurements, at simultaneous high pressures and temperatures, using a laser-heated diamond-anvil cell, on compositions of magnesium silicate perovskites that include both iron and aluminium. Attaining pressures above 1 Mbar (a million times room pressure) is fairly routine. But heating a sample at these pressures to the temperatures of the phase transition, above about 2,000 K, requires special attention. This is because diamonds are great thermal conductors and can dissipate the heat produced by the infrared lasers used to create the high temperatures, causing large temperature gradients in the sample chamber. Great care is necessary to thermally insulate the samples, which have dimensions of only tens of micrometres.

Besides the technical difficulties of reaching high pressures and high temperatures in these kinds of experiment, there is the issue of calibration. Previous results have shown a range of positive Clapeyron slopes for the transition that can be attributed, in part, to the different pressure-calibration standards used⁶. To overcome this problem, Catalli *et al.*² used the differences in the phase-boundary pressures rather than the absolute pressures: the absolute values are not important when determining the thickness of the perovskite

1. Williams, Q., Jeanloz, R., Bass, J., Svendsen, B. &

- Ahrens, T. J. Science 236, 181-182 (1987).
- Catalli, K., Shim, S.-H. & Prakapenka, V. Nature 462, 782-785 (2009).
- Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. Science 304, 855-858 (2004).
- Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Earth Planet. Sci. Lett. 224, 241–248 (2004).
- Hernlund, J. W., Thomas, C. & Tackley, P. J. Nature 434, 882–886 (2005).
- Fei, Y. et al. Proc. Natl Acad. Sci. USA 104, 9182-9186 (2007)
- Lay, T. Geophys. Res. Lett. doi:10.1029/2007gl032465 (2008).
- 8. Merkel, S. et al. Science 316, 1729-1732 (2007).
- Wookey, J. & Kendall, J. M. in Post-Perovskite: The Last Mantle Phase Transition (eds Hirose, K., Brodholt, J., Lay, T. & Yuen, D.) 171-189 (Am. Geophys. Union, 2007).
- 10. Mao, W. L. et al. Science 312, 564-565 (2006).

IMMUNOLOGY

Dendritic-cell genealogy

Sophie Laffont and Fiona Powrie

The differing origins of gut dendritic cells — white blood cells that modulate immune responses — may explain how the intestinal immune system manages to destroy harmful pathogens while tolerating beneficial bacteria.

The immune system must protect the body from invading pathogens without mounting damaging responses to its own tissues. Dendritic cells, a rare population of white blood cells, have a crucial role in determining the nature of immune reactions and in fine-tuning the balance between tolerance (where the

immune system ignores or tolerates an antigen) and the induction of inflammation to destroy pathogenic organisms. A long-standing question has been how dendritic cells drive these distinct immune outcomes. Two groups, Varol *et al.*¹ and Bogunovic *et al.*², report in *Immunity* that dendritic cells with distinct

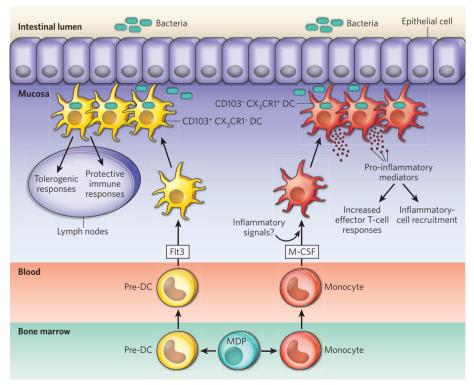


Figure 1 | Intestinal dendritic cells have different origins and different functions 12 . The two main subsets of intestinal dendritic cells (DCs) originate from distinct blood-cell precursors and depend on different growth factors for their development. Before this divergence, a common precursor, the macrophage and dendritic-cell precursor (MDP), gives rise to pre-dendritic cells (pre-DCs) and monocytes in the bone marrow. Pre-dendritic cells give rise to CD103 $^+$ CX $_3$ CR1 $^-$ dendritic cells, and depend on the growth factor Flt3 for their development, whereas monocytes develop into CD103 $^-$ CX $_3$ CR1 $^+$ dendritic cells, and depend on another growth factor, M-CSF. CD103 $^+$ dendritic cells transport microbial antigens to the lymph nodes, where they may initiate protective immune responses or promote the generation of regulatory T cells that help to maintain tolerance in the intestine. CX $_3$ CR1 $^+$ dendritic cells do not seem to migrate to lymph nodes, suggesting a more local role in promoting tissue inflammation by stimulating effector T-cell responses and producing pro-inflammatory mediators.

functions have different developmental origins, providing a cellular framework for the diverse activities of these cells.

Pioneering work by Steinman and colleagues³ in the early 1970s identified a minor population of immune cells that they named dendritic cells on the basis of their stellate shape and membranous processes. These cells were shown to be potent stimulators of another population of white blood cells, T cells. Dendritic cells are strategically placed within mucosal sites in the body, where they can detect infection and take up microbial antigens. On activation, these cells migrate to secondary lymphoid tissue, such as the lymph nodes, where they present the antigen to T cells. This activates the T cells, causing them to differentiate into effector cells that eradicate the pathogen. In the intestine, dendritic cells also promote regulatory T-cell responses that suppress immune reactions against beneficial commensal bacteria and food antigens, thereby preventing immune-related disease. Thus, intestinal dendritic cells are decision makers, ensuring selection of a T-cell response that is appropriate to the nature of the challenge to the immune system.

It is now known that dendritic cells are a diverse population of cells, differing in their anatomical location, expression of surface proteins and function. The diversity of the dendritic-cell response may reflect the differential activities of hard-wired developmentally distinct populations or may be due to different maturation states induced by environmental signals. Gaps in our knowledge of dendriticcell developmental pathways have hindered finding answers to these questions. However, recently developed genetic techniques4 to ablate dendritic-cell populations, together with an improved ability to identify specific dendritic-cell precursors, have advanced this area of research.

Dendritic cells are closely related to macrophages, which originate from white blood cells called monocytes. They are derived from a common precursor termed the macrophage and dendritic-cell precursor (MDP). Dendritic cells can also be generated⁵ from monocytes in cell culture using a growth factor called granulocyte-macrophage colony-stimulating factor (GM-CSF), and much of what we know about the biology of dendritic cells has been based on the study of these laboratory-derived cells. However, recent elegant work 6,7 has shown that monocytes and dendritic cells diverge in their developmental pathways downstream of the MDP — conventional dendritic cells in lymphoid tissue arise from a precursor cell in the blood (the pre-dendritic cell), and their differentiation depends on the growth factor Flt3. Hence, under normal conditions, blood monocytes do not give rise to dendritic cells in lymphoid tissue, raising questions about the in vivo counterpart of monocyte-derived dendritic cells.

Varol *et al.*¹ and Bogunovic *et al.*² studied the development of dendritic cells in the intestine.

These tissue dendritic cells can be identified by their expression of a surface protein, CD11c, and they can be divided further into two main subsets that express either of two surface proteins: CD103 or CX₃CR1. Both studies^{1,2} show that CD103⁺ dendritic cells follow the same developmental pathway as conventional lymphoid dendritic cells: they arise from predendritic cells without a monocyte intermediate and depend on Flt3 for their development. By contrast, both groups^{1,2} find that monocytes give rise to intestinal CX₃CR1⁺ dendritic cells (Fig. 1). Consistent with their origin from monocytes, these cells must express the macrophage growth-factor receptor known as macrophage-colony-stimulating factor (M-CSF) receptor in order to develop normally. Varol et al. also show that the GM-CSF receptor is required for monocyte-derived dendritic-cell development. However, Bogunovic and colleagues² did not observe this dependence of monocyte-derived dendritic cells on GM-CSF, perhaps reflecting the use of different markers by the two groups^{1,2} to define the dendritic-cell populations.

Why should the origins of intestinal dendritic cells differ from those of lymphoid dendritic cells? The answer may lie in the nature of the intestinal environment, which exists in a state of controlled inflammation, even in the absence of overt infection, because of continuous exposure to commensal bacteria. This controlled inflammatory state may be sufficient to recruit blood monocytes and induce their differentiation into dendritic cells locally.

Importantly, pre-dendritic-cell and monocyte-derived dendritic-cell populations in the intestine have distinct functions (Fig. 1). CD103⁺ dendritic cells migrate from tissues, and can present ingested antigen to immune cells in the local intestinal lymph nodes (mesenteric lymph nodes). Under normal conditions, these CD103⁺ cells promote intestinal tolerance by inducing the generation of regulatory T cells⁸⁻¹⁰. But CD103⁺ dendritic cells can also activate CD8⁺ killer T cells and, through production of the vitamin A metabolite retinoic acid, induce the expression of receptors on T cells that direct them to the gut¹¹. Consistent with these results, Bogunovic et al.2 show that CD103+ dendritic cells take up salmonella bacteria that have invaded the intestine and then transport the bacteria to mesenteric lymph nodes. Taken together, these data identify pre-dendritic-cell-derived CD103⁺ dendritic cells as important mediators of immune surveillance in the intestine that can promote both host-protective and tolerogenic T-cell responses.

Monocyte-derived CX₃CR1⁺ dendritic cells, on the other hand, have been associated with the induction of inflammatory T cells that promote intestinal inflammation¹². CX₃CR1⁺ dendritic cells have been shown to extend processes into the intestinal lumen to sample antigen. But Bogunovic *et al.*² provide evidence that these

cells do not usually migrate to the mesenteric lymph nodes. Furthermore, Varol $et\ al.^1$ demonstrate that CX_3CR1^+ dendritic cells alone are sufficient to drive intestinal inflammation by producing the pro-inflammatory mediator tumor necrosis factor- α . These observations raise the possibility that intestinal monocytederived dendritic cells do not initiate T-cell responses in secondary lymphoid tissue, but rather promote the inflammatory response at the site of pathogen entry. Further studies are required to assess the behaviour of this subset of dendritic cells during intestinal infection.

The identification of the different progenitors of intestinal dendritic cells and the growth factors that control their development are crucial first steps to investigating the functional role of these distinct cell populations in health and disease. This important conceptual advance also potentially opens the door for targeted manipulation of dendritic-cell subsets to treat immune-mediated diseases.

Sophie Laffont and Fiona Powrie are in the
Translational Gastroenterology Unit, Nuffield
Department of Clinical Medicine, John Radcliffe
Hospital, University of Oxford, Oxford OX3 9DU,
and at the Sir William Dunn School of Pathology,
Oxford, UK.

e-mail: fiona.powrie@path.ox.ac.uk

- 1. Varol, C. et al. Immunity 31, 502-512 (2009).
- 2. Bogunovic, M. et al. Immunity 31, 513-525 (2009).
- Steinman, R. M. & Cohn, Z. A. J. Exp. Med. 137, 1142–1162 (1973).
- 4. Jung, S. et al. Immunity 17, 211-220 (2002).
- Lutz, M. B. et al. J. Immunol. Methods 223, 77–92 (1999).
- Waskow, C. et al. Nature Immunol. 9, 676-683 (2008).
- 7. Liu, K. et al. Science 324, 392-397 (2009).
- 8. Coombes, J. L. et al. J. Exp. Med. 204, 1757-1764 (2007).
- 9. Sun, C. M. et al. J. Exp. Med. 204, 1775-1785 (2007).
- 10. Jaensson, E. et al. J. Exp. Med. 205, 2139-2149 (2008).
- Johansson-Lindbom, B. et al. J. Exp. Med. 202, 1063–1073 (2005).
- 12. Atarashi, K. et al. Nature 455, 808-812 (2008).

STRUCTURAL BIOLOGY

Molecular coin slots for urea

Mark A. Knepper and Joseph A. Mindell

Membrane-bound protein channels that allow only urea to pass through are vital to the kidney's ability to conserve water. Crystal structures show that the channels select urea molecules by passing them through thin slots.

Coin-operated vending machines must reliably accept only valid coins of the correct denomination. Modern machines recognize coins on the basis of their size, shape and even their chemical composition (determined by measuring the coins' electromagnetic properties). On page 757 of this issue, Levin et al.1 describe the crystal structure of a bacterial urea-conducting channel that acts like a molecular version of a coin-operated machine — it selectively allows planar urea molecules to pass through on the basis of their size, shape and electrical-charge distribution. The authors' results thereby provide insight into the function of a class of channel that is vital to the function of the human kidney.

Urea is a small, nitrogen-containing organic molecule. Although uncharged, it is highly polar and adept at hydrogen bonding; indeed, urea's ability to form hydrogen bonds with similarly polar water molecules accounts for its astonishingly high solubility in water (more than 8 M). Urea has a special place in the history of science, because it was the first organic molecule to be synthesized by non-biological means from inorganic starting materials². It is also a key molecule in human physiology: in the liver, excess nitrogen from the normal breakdown of proteins is incorporated into urea, which is subsequently released into the bloodstream, ultimately to be excreted by the kidneys. Because of its high solubility and low toxicity, large amounts of urea can be excreted in small amounts of water, allowing the kidney to conserve water even when urea excretion is high.

But urea excretion presents a challenge for the kidney by virtue of the osmotic forces that the compound generates. The basic structural and functional unit of the kidney is the nephron. Each nephron filters water and small molecules (including urea) from the blood, creating a flow of fluid that passes through a long, narrow renal tubule. These tubules modify the filtered fluid using myriad transport processes, and what is left becomes urine. The final part of the renal tubule is called the collecting duct. Left uncontrolled, the osmotic force of the highly concentrated urea in collecting ducts would suck water from the kidney interstitium (the space between renal tubules), thus undesirably increasing water excretion, a process called osmotic diuresis (Fig. 1a, overleaf). To avoid this, the kidney must balance urea concentrations inside and outside the urinary space by allowing urea to move rapidly across the membranes of the epithelial cells that line the collecting duct (Fig. 1b).

The polarity of urea molecules prevents them from readily penetrating nonpolar lipid membranes. Kidney cells therefore use specialized channel proteins to move the molecule rapidly into and through cells. These channels — called UT-A1 and UT-A3 — allow urea

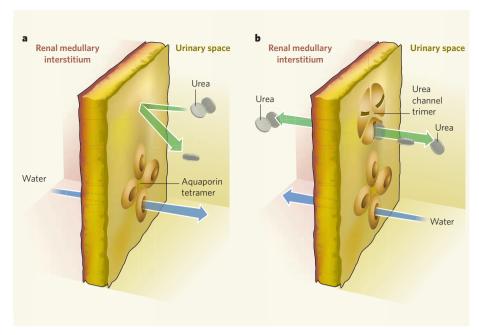


Figure 1 | Urea channels allow the kidney to conserve water. Urine, formed in the kidney, contains high concentrations of urea. Urine passes through the interior urinary space of the collecting ducts (the final components of the kidney) before being excreted. a, In the absence of urea channels, the high concentration of urea in the urinary space would create an osmotic pressure that would draw water from the renal medullary interstitium (the space between the renal tubules) through aquaporin channels. This process, known as osmotic diuresis, would be associated with a high rate of water excretion from the kidney. b, In normal kidney, urea channels in the duct system's cell membranes allow urea to equilibrate between the urinary space and the interstitium, so that no osmotic pressure difference occurs because of urea. Salts such as sodium chloride and potassium chloride accumulate in the interstitium independently of urea, and create the osmotic force that drives water out of the urinary space through aquaporins. Levin et al. 1 report the crystal structure of a bacterial urea channel, and find that the protein uses slot-shaped pores to recognize and pass flat urea molecules, like the coins in a vending machine.

concentrations to equilibrate rapidly between the urine and interstitium. As a result, urea accrues to extremely high concentrations in the interstitium, osmotically balancing the large amount of urea in the urine and making the urinary urea osmotically 'invisible' — thereby allowing the kidney to conserve water. Consistent with this posited role of urea channels, genetically modified mice lacking the gene that encodes collecting-duct urea channels cannot efficiently conserve water unless fed a low-protein diet to reduce urinary urea³. Inhibitors of urea channels have therefore been proposed as a new class of drug for treating clinical disorders that involve water retention⁴.

The channel characterized by Levin *et al.*¹, dvUT, is a bacterial protein whose amino-acid sequence resembles that of the kidney urea channels. Because these bacterial proteins don't necessarily have the same function as their mammalian cousins, the authors first tested whether dvUT can carry urea. Gratifyingly, they found that, when expressed in cell membranes, dvUT does allow urea to permeate cells. Furthermore, the authors could inhibit urea movement using phloretin, a blocker of mammalian urea channels.

Levin and colleagues' crystal structure¹ of dvUT reveals that it is similar to other proteins that move small neutral molecules across

membranes, such as aquaporins (water channels)⁵ and amtB (an ammonia channel)⁶. Like those proteins, dvUT forms an oligomeric complex in the membrane (a trimer, in this case). Also like those proteins, each monomer in the dvUT assembly seems to contain an independent pathway for the movement of its substrate. Completing the family resemblance, all three classes of channel have an 'inverted repeat' motif, in which both halves of the molecule have similar amino-acid sequences but adopt opposite orientations with respect to the plane of the membrane.

The crystal structure¹ also reveals much about the mechanism of dvUT action. First, there is an unoccluded pathway for urea to travel through the protein. Although urea itself could not be visualized when incorporated into the crystallizing protein, when the authors incorporated a urea analogue, dimethylurea, they observed that it binds at two sites within this pathway. The presence of a clear conduit is the hallmark of proteins that operate by a channel mechanism, in which substrates pass through without requiring extensive changes to the conformation of the protein. Many in the field had assumed that a 'carrier' model — which typically involves large conformational changes - would be needed to explain the kinetics of urea transport⁷. Other evidence, however, had already suggested channel-like behaviour^{8,9}, as strongly implied

by Levin and colleagues' structure¹.

The second insight obtained from Levin and colleagues' work is the 'molecular coin-slot' mechanism that allows urea to pass through dvUT in preference to other small molecules. At the entrance to the protein's pore, the planar, aromatic side chains of two phenylalanine amino-acid residues form a hydrophobic slot just wide enough to permit the coin-shaped urea molecule to enter (see Fig. 4 on page 760). This fixes the orientation of the urea molecules, setting them up to be grabbed by a line of oxygen atoms — dubbed the "oxygen ladder" by the authors — lying along one side of the slot. The oxygen ladder provides electrostatic interactions for the urea molecules, helping to extract them from the water environment outside the channel.

The whole dvUT pore actually consists of two slots in series, separated by a gap that is lined by hydrophobic amino-acid side chains. The authors' crystal structure of a dvUTdimethylurea complex reveals two dimethylurea molecules, each one bound to the oxygen ladders at opposite ends of the pore and snuggled between phenylalanine residues at the pore's entrance. The molecules seem to be stabilized in these positions by interactions between their delocalized electrons and the aromatic electrons in the phenylalanine side chains. Thus, the molecular coin slot of dvUT might detect not only urea's size and planar shape, but also its unique electronic configuration.

Levin and colleagues' study' opens the door to further investigation of the properties of this family of urea channels. This should shed light on such issues as the oligomeric organization of mammalian urea channels; the energetics of the competition between the pore's lining and water outside the channel for binding to urea; and the regulation of the channels by gating or trafficking mechanisms. Such work will provide a treasure trove of information — you can put money on it.

Mark A. Knepper is in the Epithelial Systems Biology Laboratory, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-1603, USA. Joseph A. Mindell is at the Membrane Transport Biophysics Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892-4066, USA. e-mail: knep@helix.nih.gov

- Levin, E. J., Quick, M. & Zhou, M. Nature 462, 757-761 (2009)
- 2. Wöhler, F. Ann. Phys. Chem. (Leipz.) **88,** 253–256 (1828).
- Fenton, R. A., Chou, C. L., Stewart, G. S., Smith, C. P. & Knepper, M. A. Proc. Natl Acad. Sci. USA 101, 7469–7474 (2004).
- Levin, M. H., de la Fuente, R. & Verkman, A. S. FASEB J. 21, 551–563 (2007).
- 5. Murata, K. et al. Nature 407, 599-605 (2000).
- 6. Khademi, S. et al. Science **305**, 1587-1594 (2004).
- Mayrand, R. R. & Levitt, D. G. J. Gen. Physiol. 81, 221–237 (1983).
- Levine, S. D. & Worthington, R. E. J. Membr. Biol. 26, 91-107 (1976).
- Finkelstein, A. Water Movement Through Lipid Bilayers, Pores and Plasma Membranes: Theory and Reality (Wiley, 1987).

OBITUARY

Qian Xuesen (1911-2009)

Founder of China's missile and space programme.

Perhaps no one better embodies the irony of the cold-war era than Qian Xuesen, also known as Tsien Hsue-shen, the Chinese rocket scientist who died on 31 October at the age of 97. Widely acknowledged for his work on rocket propulsion and his contribution to the US missile programme during the Second World War, Qian was deported at the height of the McCarthyism hysteria on dubious charges of being a communist, only to become the driving force behind China's rise to the first rank of space nations.

Qian was born in the eastern city of Hangzhou in 1911, as China's 2,000-yearold feudal system was overthrown, and he was destined to encounter both turbulent historic events and social upheaval. After graduating from Shanghai Jiao Tong University as an engineer, in 1935 he went to study aeronautics at the Massachusetts Institute of Technology (MIT) in Cambridge. Theoretically inclined, he was dissatisfied with the practical emphasis that MIT placed on its curricula at the time, and soon moved to the California Institute of Technology (Caltech) in Pasadena. During his time there, he rose to become one of the most prominent rocket scientists in the United States.

At Caltech, Qian was the protégé of the legendary theoretical aerodynamicist Theodore von Kármán, and became a renowned theoretician in high-speed flight. The two men were the leading members of a group of rocket experimenters known as the Suicide Squad because of the dangerous nature of their work. In 1939, their research attracted the attention of the US Army Air Corps, which tasked Caltech, including Qian, with developing jet-assisted take-off technology — a means of helping heavy aircraft to take off by providing additional thrust in the form of small rockets.

The 1943 discovery of German rocketry development led to the acceleration of the US missile programme and the creation of the Jet Propulsion Laboratory at Caltech, which appointed Qian as a director. Two years later, Qian received high-grade security clearance from the Pentagon and, as a member of the Scientific Advisory Board, began advising on the latest classified technologies for military development. By 1949, he had laid the theoretical groundwork for a spaceplane with winged rockets, a predecessor of the space shuttle.

Meanwhile, Sino-US relations had deteriorated with the onset of the cold war, and this had devastating consequences for Qian's career. In 1950, he was accused of



being a communist spy — allegations that have never been substantiated — and his security clearance was revoked. Deeply hurt by the accusations, and realizing that he was no longer welcome in the United States, Qian attempted to return to China, whereupon he was arrested and imprisoned by the US Immigration Service for two weeks. He spent the next five years living under partial house arrest, enduring constant humiliation and harassment, although he was able to continue teaching and to conduct limited research at Caltech. He was eventually allowed to leave the United States in 1955, as part of the conditions of China's release of eleven American soldiers captured during the Korean War.

Qian received a hero's welcome in China, and was soon asked to create a missile programme — a daunting challenge in a country where even telephones were luxuries. He began by retraining Chinese scientists and engineers, attracting Western-trained Chinese researchers to return, and building an efficient education and management system for research and development, all of which were to have lasting effects on the country's rocketry development. He joined the Communist Party in 1958 and became a trusted, high-ranking party official. With access to the top Chinese leaders, including Chairman Mao Zedong, he was able to persuade officials to support whatever measures he felt China needed to progress.

In 1956, Qian founded the Institute of Mechanics in Beijing, now one of the world's leading aeronautics institutions, and influenced the teaching of engineering at many Chinese universities. Under his leadership, China progressed from copying the intermediate-range Soviet R-2 missile (a version of the V-2 rocket developed by Nazi Germany) to launching Dongfeng-2 in 1964 (a medium-range ballistic missile based on home-grown technology), and successfully tested its first atomic bomb a few months later. China's first nuclear ballistic missile was launched using Dongfeng-2 two years later, and the first Chinese satellite went into orbit using Long March-1, a three-stage space-launch version of Dongfeng-4, in 1970.

Qian's endeavours were responsible for the development of the Haiying family of anti-ship missiles, widely known in the West as Silkworms. They also paved the way for a succession of increasingly ambitious space programmes. For example, Shenzhou-5, China's first manned space mission, was launched in 2003, and Chang'e-1, the first Chinese lunar probe, was sent into orbit in 2007, using the Long March family of rockets.

It is extraordinary that China's missile and space programmes took off against a backdrop of economic and political turmoil. They were postponed several times during the Great Leap Forward — a social and economic plan enacted in 1958 that brought the country's economy to a standstill and triggered widespread famine. In an ironic twist of fate, Qian also endured the Cultural Revolution, an 'anti-revolutionary' witch-hunt launched by Mao in 1966. Some of Qian's friends and colleagues were removed from their positions as a result; Qian himself, although largely unaffected, was forced to write letters of 'confession'.

Despite the grave injustice Qian suffered in the United States, it is clear that his time at Caltech was one of the most enjoyable periods of his life. He revelled in the great relationship he had with von Kármán, enjoying his mentor's jokes and their often heated arguments. He also reminisced, with great affection, about the intellectual ethos and creative spirit fostered at Caltech. He contrasted these with China's academic culture and science infrastructure, which he thought had not, and would not, lead to any real scientific innovation.

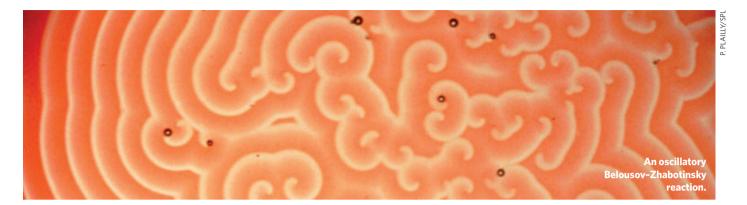
Towards the end of his life, Qian was confined to bed, but he still received regular visits from the Chinese premier Wen Jiabao, with whom he discussed the problems that he felt affected China's scientific development. Although delighted to witness the tremendous advances in the missile and space programmes that he had helped to establish, Qian felt that the intellectual legacy he had brought from Caltech had largely failed to take root in the academic soil of his own country, and this saddened him deeply. Despite his reservations, there is no doubt that he sowed the seeds of change for Chinese science.

Jane Qiu

Jane Qiu writes for *Nature* from Beijing. e-mail: jane@janeqiu.com

Nature Vol 462|10 December 2009





SYSTEMS CHEMISTRY

Molecular networks come of age

Jonathan R. Nitschke

The advent of sophisticated analytical tools enables the collective behaviour of networks of interacting molecules to be studied. The emerging field of systems chemistry promises to allow such networks to be designed to perform complex functions, and might even shed light on the origins of life.

What is systems chemistry?

It's the study of complex systems, or networks, of molecules. Tools for analysing complex networks are being developed and employed in fields as diverse as computer science and sociology. By applying these tools to systems of interacting molecules — molecules that might link together into larger superstructures, or catalyse one another's formation — chemists can investigate how interactions between members propagate through networks, allowing complex behaviour to emerge.

So what questions are systems chemists asking?

There are two main questions. The first is how the complex networks of molecules found on the prebiotic Earth might have crossed the threshold of life. Research into this area seeks possible mechanisms for how the biochemical building blocks of life were selected, and how biomolecules developed to have only one 'handedness' (chirality). The second question is how collections of molecules self-assemble into complex structures, and how secondary interactions between molecules and competition for molecular building blocks lead to complex behaviour within self-assembling systems. Such systems of molecules might sort themselves, for example, by pairing up in the same way as complementary DNA bases, or they might adapt to the addition of a substrate by combining in such a way as to create a strong binding pocket for that substrate.

Why the increased interest just now?

Recent strides in analytical methods allow complex mixtures of compounds to be picked apart and their constituents quantified without isolating each one individually. Using techniques such as high-performance liquid chromatography coupled to mass spectrometry and multidimensional nuclear magnetic resonance, up to 1,000 or so individual compounds can be identified within a mixture. The chemical evolution of mixtures over time can be tracked, allowing complex transformations to be followed and kinetic relationships between network members to be deciphered. Discoveries, ideas and tools from systems biology have also informed and inspired work in systems chemistry. Chemists can help to define the bedrock on which systems biology is built by showing how interactions between molecules lead to the emergence of function.

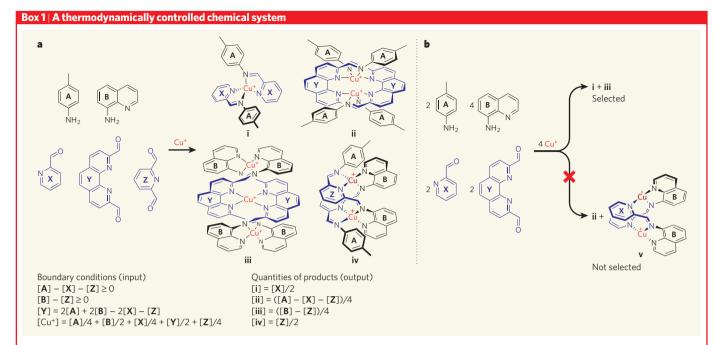
What do you mean by 'emergence'?

Emergence occurs when a complex system exhibits properties that can't be predicted by considering its subcomponents in isolation. For example, consciousness emerges when neurons come together into a brain, and the recent global recession was a consequence of vast numbers of interwoven individual financial transactions. Predictability is subjective, however, so calling a property 'emergent' is uncomfortably close to saying 'I wasn't clever enough to predict it', or worse — 'it is mystical'. Perhaps a better definition of 'emergent' is

'interesting and counter-intuitive', which highlights the necessary subjectivity and ties the elucidation of emergent phenomena directly to the scientist's role in unravelling the predictive rules that underlie everyday phenomena.

And what do you mean by 'complexity'?

In systems chemistry, perhaps the most useful definition is borrowed from information science: complex systems require more information to describe them than do simple systems. Specifically, complex chemical systems are richer in connections between members than simple ones. The complexity of a single molecule is necessarily limited by the number of bonds that individual atoms may form with others. But the complexity of a collection of molecules can be more than the sum of the complexities of each individual molecule, if the molecules can interact in different ways. Such interactions might be supramolecular, with molecules fitting together like locks and keys; or they might involve the exchange of molecular subcomponents; or they could be catalytic if one molecule catalyses the formation of several others. It becomes easier for complex behaviour to emerge from networks as the connections between members grow denser, because networks can then respond to stimuli in more complicated ways — the addition of a single chemical signal to a system can affect many of its members through the transmission of the signal from one member to another.



In thermodynamically controlled systems, networks of reversible reactions lead to the generation of the most stable set of products. a, In this system, amines A and B react reversibly with aldehydes X, Y and Z and with copper ions to form copper complexes that contain imine (C=N) groups. In

principle, a large number of products could form. But in practice, two empirical self-assembly rules are observed: the system generates the smallest possible structures that are not unduly strained, and the products are all coordinatively saturated (all the copper ions are bound to four imines, and each

imine is bound to a copper ion). Only structures **i-iv** satisfy these rules. **b**, Additional selection rules arise in cases where several products form simultaneously. In this system, for example, **Y** preferentially pairs with **B** to make **iii**, so that **A** and **X** are left to generate **i**. Even though **A** could pair with **Y** to make **ii**, **B** and **X** would

then be forced to make **v**, which is not coordinatively saturated. The system may thus be seen to carry out logic operations during self-assembly, described by the boundary conditions and equations shown. [A] represents the concentration of **A**, [B] that of **B**, and so on.

J.R.N.

What kinds of chemical system exhibit complex behaviour?

There are two main categories. The first involves systems of molecules undergoing thermodynamically controlled reactions (reversible reactions that lead to the generation of the most stable set of products), which approach equilibrium in a complex way. In these systems, numerous interactions between the molecules may act together to determine the most stable set of products from the network as a whole. The second category involves sets of kinetically controlled reactions (irreversible reactions proceeding along the most expeditious paths available), in which one reaction's products modulate another reaction's rate within the system.

How can an approach to thermodynamic equilibrium be complex?

When there are several different ways in which a pool of building blocks can come together to form a collection of products, the formation of one structure may leave behind a set of building blocks that can't form a stable product. The collective stability of all product structures within a system must thus be considered, which can be a complicated undertaking. The rules that govern the behaviour of a self-sorting system may be traced through and expressed in the form of algorithms or equations (Box 1).

What kinds of complex system evolve under kinetic control?

Those that contain autocatalytic reactions, in which a product catalyses its own formation, and those that involve cross-catalytic reactions, in which one product catalyses another's formation. Cross-catalytic reactions can give rise to oscillations, as a system periodically switches between discrete states (Box 2, overleaf). Feedback between coupled oscillators can be used to carry out computations, as happens in our own neural circuitry.

Do any systems exhibit both thermodynamic and kinetic complexity?

Biological systems do. Complex self-assembly processes operating under thermodynamic control are used to construct many of the complex forms of biomolecular systems. The protein subunits of viral capsids (their outer shells), for example, are held together in highly symmetric assemblies by many discrete, reversible interactions. In many cases, viral genetic material serves as a template for capsid formation — negatively charged nucleic acids attract positively charged capsid-forming proteins, which then arrange themselves into a capsid through the thermodynamic equilibration of van der Waals attractions, hydrophobic effects and hydrogen bonds, among other forces. Alongside such thermodynamically controlled processes, complex kinetics underlies all

biological signalling networks. The simplest biological systems are therefore much more complicated than the most intricate synthetic chemical systems currently known. This begs the question of which chemical systems on the prebiotic Earth might have developed sufficient complexity to become alive.

Do we know the answer?

Not yet, but ideas abound — such as the hypothesis that self-replicating RNA-like molecules, able both to encode information and to act as catalysts, served as an intermediate stage between naturally occurring mixtures of complex organic molecules and living systems. But we cannot yet follow our own story back to its beginning. This is a primal quest for chemists, and one that is likely to spin off much new knowledge along the way.

Can chemical systems perform complicated tasks?

Definitely. For example, a molecular 'motor' has been made in which a rotor spins unidirectionally about a double bond when irradiated with ultraviolet light (Fig. 1, overleaf). When the motor is dissolved in a thin film of liquid-crystal (LC) matrix, molecular interactions between the turning rotor and the matrix lead to a continuous reorganization of the overall LC system. This causes micrometre-scale corrugations on the LC surface to

Box 2 | Kinetically controlled chemical systems

Time (t)

 $d[\mathbf{D}]/dt = k_1[\mathbf{D}][\mathbf{C}] - k_2[\mathbf{D}][\mathbf{E}]$ $d[\mathbf{E}]/dt = k_3[\mathbf{E}][\mathbf{D}] - k_A[\mathbf{E}]$

Kinetically controlled complex chemical systems contain autocatalytic reactions, in which a product catalyses its own formation, and cross-catalytic reactions, in which one product catalyses another's formation. The graph shows the changes in

concentration of the components of a hypothetical chemical system containing three compounds — \mathbf{C} , \mathbf{D} and \mathbf{E} — described by the rate equations under the graph. [\mathbf{C}] indicates the concentration of \mathbf{C} , [\mathbf{D}] that of \mathbf{D} , and so on; t is time; k_1 to k_4 are constants. The

system is both autocatalytic, because the rate of increase of [E] is proportional to [E], and cross-catalytic, because the rate of increase of [E] is proportional to [D]. As long as [C] is kept constant, the system will oscillate regularly.

Ecologists will recognize the equations as the Lotka-Volterra 'predator-prey' equations of population dynamics in ecosystems, where $\bf D$ is prey, $\bf E$ is predator, $\bf C$ is food and the constants reflect rates of prey reproduction (k_1) , predation (k_2) , predator reproduction (k_3) and predator mortality (k_4) . The polymath Alfred Lotka originally conceived the equations to describe chemical oscillations; only later were they applied to

ecosystems. No known chemical reactions follow these equations, but slightly more complex autocatalytic and cross-catalytic systems, such as the oscillatory Belousov–Zhabotinsky reaction, are described by similar mathematics.

The myriad auto- and cross-catalytic biochemical reactions that take place during the cell cycle may be seen as constituents of a highly complex chemical oscillation in which an entire cell is duplicated. A fundamental link thus connects the kinetics of the biochemical reactions underpinning the exponential growth of cells (and by extension, organisms) to the dynamics of populations of organisms within ecosystems.

J.R.N.

rotate clockwise, which in turn can rotate a macroscopic object embedded in the matrix. Interactions between the components of the system thus cooperate to transform light energy into torque. Light has also been used to push a different chemical system out of equilibrium, by allowing ring-shaped molecules to travel easily in one direction along molecular shafts, but creating a dynamic barrier to their return. This system is described as an information ratchet by its creators, because it can process information about the ring's position, and the barrier can be raised or lowered accordingly. Such a mechanism could form the basis of a component for information-processing molecular machines.

What other problems might be addressed using systems chemistry?

Many advanced materials, such as composites, consist of systems of intermingled basic materials (metals, ceramics, polymers and so on) that interact in complex ways. Understanding the chemical nature of these interactions is essential for tuning and designing the properties of advanced materials. Synergistic drug interactions — interplay between different drugs in the body that can cause biological effects not seen when the drugs are taken individually — likewise emerge from complex networks of drug molecules and biomolecules. Insight into such networks will underpin our understanding of the pharmacology of drug combinations.

Biological structures such as the ribosome (nature's protein-making machinery) also provide clues to how chemical systems might be used as molecular assembly lines to effect multi-step transformations, as a substrate is passed from one component of a catalytic system to another, with each component bolting on a new molecular subunit.

What will be needed to meet these goals?

Perhaps the most useful tool would be a 'retrosynthetic' methodology that can be used to reverse-engineer chemical systems. This would facilitate the design of links between the members of chemical networks, in the same way that sequences of chemical bonds may be planned and constructed in the syntheses of complex organic molecules. Such a methodology for systems chemistry is still a long way off, but its foundations are being laid as the basic rules that govern interactions between members of molecular networks are deciphered.

Jonathan R. Nitschke is in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. e-mail: jrn34@cam.ac.uk

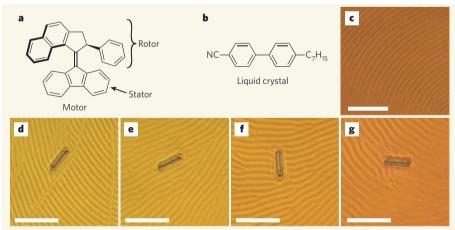


Figure 1 | Systems chemistry in a spin. a, When this molecular 'motor' is irradiated with ultraviolet light, the rotor turns relative to the stator. b, c, If the motor molecules are dissolved in a thin film of a liquid crystal (LC; b) at a concentration of 1% by weight, a 'cholesteric' phase (c) is created, in which the LC molecules stack in a helical fashion, parallel to the surface they cover. This results in a sinusoidal surface corrugation of the film, with a peak-to-peak distance that corresponds to the helices' pitch. Light-driven twisting of the motor within the LC matrix lengthens the optimal helical pitch of the cholesteric LC matrix, 'winding' tension into the system. d-g, As the LC molecules rearrange to release this tension, the surface corrugations rotate clockwise. The resulting torque is sufficient to rotate a cylindrical glass rod (5×28 micrometres in size), as shown in this series of pictures taken at intervals of 15 seconds. The transformation of light energy into torque thus emerges out of a complex system formed by two simple molecules. Scale bars, 50 micrometres. (Images taken from R. Eelkema et al. Nature 440, 163; 2006.)

FURTHER READING

523-527 (2007).

Ludlow, R. F. & Otto, S. Systems chemistry. *Chem. Soc. Rev.* **37,** 101–108 (2008).

Vidonne, A. & Philp, D. Making molecules make themselves — the chemistry of artificial replicators. *Eur. J. Org. Chem.* 593–610 (2009).

Sarma, R. J. & Nitschke, J. R. Self-assembly in systems of subcomponents: Simple rules, subtle consequences. *Angew. Chem. Int. Edn* 47, 377–380 (2008).

Dadon, Z., Wagner, N. & Ashkenasy, G. The road to non-enzymatic molecular networks. *Angew. Chem. Int. Edn* 47, 6128–6136 (2008).

Eelkema, R. et al. Molecular machines: Nanomotor rotates microscale objects. *Nature* **440**, 163 (2006). Serreli, V., Lee, C.-F., Kay, E. R. & Leigh, D. A. A molecular information ratchet. *Nature* **445**,

ARTICLES

Cancer-associated IDH1 mutations produce 2-hydroxyglutarate

Lenny Dang¹, David W. White¹, Stefan Gross¹, Bryson D. Bennett², Mark A. Bittinger¹, Edward M. Driggers¹, Valeria R. Fantin¹, Hyun Gyung Jang¹, Shengfang Jin¹, Marie C. Keenan¹, Kevin M. Marks¹, Robert M. Prins³, Patrick S. Ward⁴, Katharine E. Yen¹, Linda M. Liau³, Joshua D. Rabinowitz², Lewis C. Cantley⁵, Craig B. Thompson⁴, Matthew G. Vander Heiden¹† & Shinsan M. Su¹

Mutations in the enzyme cytosolic isocitrate dehydrogenase 1 (IDH1) are a common feature of a major subset of primary human brain cancers. These mutations occur at a single amino acid residue of the IDH1 active site, resulting in loss of the enzyme's ability to catalyse conversion of isocitrate to α -ketoglutarate. However, only a single copy of the gene is mutated in tumours, raising the possibility that the mutations do not result in a simple loss of function. Here we show that cancer-associated IDH1 mutations result in a new ability of the enzyme to catalyse the NADPH-dependent reduction of α -ketoglutarate to R(-)-2-hydroxyglutarate (2HG). Structural studies demonstrate that when arginine 132 is mutated to histidine, residues in the active site are shifted to produce structural changes consistent with reduced oxidative decarboxylation of isocitrate and acquisition of the ability to convert α -ketoglutarate to 2HG. Excess accumulation of 2HG has been shown to lead to an elevated risk of malignant brain tumours in patients with inborn errors of 2HG metabolism. Similarly, in human malignant gliomas harbouring IDH1 mutations, we find markedly elevated levels of 2HG. These data demonstrate that the IDH1 mutations result in production of the onco-metabolite 2HG, and indicate that the excess 2HG which accumulates *in vivo* contributes to the formation and malignant progression of gliomas.

Mutations in the enzyme cytosolic isocitrate dehydrogenase 1 (IDH1) are found in approximately 80% of grade II-III gliomas and secondary glioblastomas in humans¹⁻³. These mutations occur at a single amino acid residue of IDH1, arginine 132, which is most commonly mutated to histidine (R132H)^{1,3,4}. Only a single copy of the gene has been found to be mutated in tumours¹⁻⁶. Many of the high-grade gliomas with IDH1 mutations are secondary glioblastomas that have progressed from lower grade lesions^{1–3,5}. When analysed in relation to other genes implicated in brain tumours, the compiled evidence suggests that *IDH1* is often the first mutation that occurs². Although these findings suggest that IDH1 mutations are selected for early during tumorigenesis, why mutations in a single allele of IDH1 result in predilection for malignant progression is uncertain. It has been reported that the R132H mutation disrupts the ability of IDH1 to convert isocitrate to α-ketoglutarate^{3,7}, but the consequences of this impaired enzymatic activity on cellular metabolism have not been systematically analysed. For example, although R132 IDH1 mutations might reduce the rate of cytosolic α-ketoglutarate production as suggested by others⁷, whether IDH1 mutations can influence the enzyme's ability to act on α ketoglutarate as a substrate has not been explored. This latter activity may be particularly important for the tumorigenic role of IDH1 mutations because cytosolic α-ketoglutarate is in equilibrium via transamination with glutamate that has a unique role in glia cell physiology, and IDH1 mutations are especially prevalent in malignant gliomas.

IDH1 mutant expressing cells have elevated 2HG levels

To understand the impact of IDH1 mutation on cellular metabolism, we profiled metabolites to identify changes in metabolite levels in

cells expressing R132 mutant IDH1 compared with cells expressing wild-type IDH1. To initiate these studies, we stably transfected U87MG glioblastoma cells, which are wild-type for IDH1, with Myc-tagged wild-type or R132H mutant IDH1. Cells expressing either Myc-tagged wild-type or mutant IDH1 were used for metabolite profiling experiments (Fig. 1a). Metabolites extracted from exponentially growing cells were profiled by liquid chromatography-electrospray ionization-mass spectrometry (LC-MS). In initial survey analyses, full-scan LC-MS in negative ion mode (exact mass) was used to examine differences in metabolite species with an m/z between 110 and 1,000. Relative quantitative data were collected for approximately 850 ions and identities were proposed by comparison with known human metabolites. Identities of >100 species identified by a combination of exact mass and retention time match to purified standards were assigned⁸. There were no significant differences between cells expressing wild-type IDH1 when compared with parental cells. The levels of most observed ions were also similar between wild-type and R132H mutant IDH1 expressing cells (Fig. 1b), with no significant changes found in canonical tricarboxylic acid (TCA) cycle species (P > 0.05 for citrate, isocitrate, α -ketoglutarate, succinate, fumarate and malate). However, three species were significantly more abundant in R132H mutant IDH1 expressing cells (P < 0.001 for each). The mass of one of these ions matched precisely to 2HG (expected m/z147.0299; measured 147.0299). The other ions co-eluted with 2HG, and had masses consistent with the sodium adduct and a dehydrated form of 2HG. Subsequent injection of 2HG standard confirmed a retention time match to the biological peak, and that it forms all three of the observed ions during LC–MS ionization (data not shown). The cellular accumulation of 2HG was quantified by targeted

¹Agios Pharmaceuticals, Cambridge, Massachusetts 02139, USA. ²Department of Chemistry and Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. ³Department of Neurosurgery, UCLA Medical School, Los Angeles, California 90095, USA. ⁴Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁵Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA. †Present address: Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

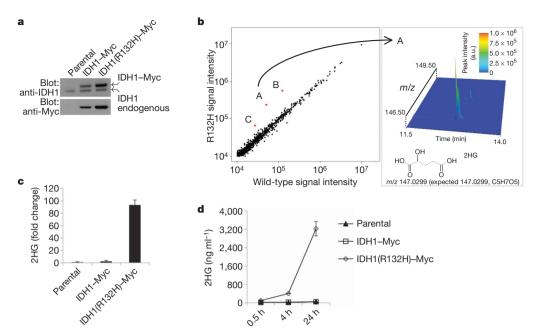


Figure 1 | Cells expressing human R132H IDH1 contain markedly elevated levels of 2HG. a, Western blots for Myc-tagged human isocitrate dehydrogenase 1 (IDH1–Myc) or R132H mutant (IDH1(R132H)–Myc) in stably transfected U87MG human glioblastoma cells. b, Metabolite profiles from cells expressing R132H IDH1 or wild-type IDH1 detected by LC–MS scanning for species between 110–1,000 m/z (M-H $^+$). Red spots labelled A, B and C represent species assigned to 2HG, dehydro-2HG and 2HG-sodium adduct, respectively. Spectrometric details supporting the identification of

species 'A' as 2HG are shown in the right panel. **c**, Cells expressing R132H IDH1 contain elevated levels of 2HG. Data were normalized by cell number and expressed as fold difference relative to parental values. Error bars depict one standard deviation (s.d.) from the mean of three independent experiments. **d**, Cells expressing R132H IDH1 display time-dependent accumulation of 2HG in cell culture media, normalization was as described in **c**. Errors bars depict one s.d. from the mean of four independent experiments.

triple-quadrupole LC–MS–MS analysis of cell extracts (Fig. 1c). The structure of 2HG is close to α -ketoglutarate, the product of the IDH1 enzyme. Thus, the sole metabolite identified by untargeted metabolite profiling to be markedly altered by R132H mutant IDH1 expression was also implicated by its structure to be IDH1-related.

The accumulation of 2HG was not restricted to cell extracts, as 2HG was found to accumulate rapidly in the medium of cells expressing R132H mutant IDH1 (Fig. 1d). No appreciable 2HG could be found in the medium of wild-type cells or cells transfected with wildtype IDH1 (Fig. 1d). Isotope-labelling experiments on whole cells using uniformly labelled ¹³C-glutamine as a culture media nutrient demonstrated that the carbons in 2HG are derived from glutamine, with reasonably high overall pathway flux from glutamine through glutamate and α-ketoglutarate to 2HG. Moreover, labelling experiments did not demonstrate any other major alterations in central carbon metabolic flux in cells expressing R132H mutant 2HG (data not shown). The presence of a Myc epitope tag did not alter activity of R132H mutant IDH1. Despite being expressed at lower levels than the Myc-tagged R132H IDH1, cells transfected with untagged R132H IDH1 demonstrated a comparable increase in 2HG production (Supplementary Fig. 1). To determine whether 2HG production in cells expressing R132H mutant IDH1 is unique to U87MG cells, we stably expressed wild-type and R132H mutant IDH1 in wild-type IDH1-expressing LN-18 glioblastoma cells (Supplementary Fig. 2a). Similar to results obtained with U87MG cells, the major difference in metabolite levels observed in LN-18 cells expressing R132H mutant IDH1 was an increased level of 2HG (Supplementary Fig. 2b).

Mutant IDH1 directly converts α -ketoglutarate to 2HG

The R132H mutation has been reported to result in loss of function for enzyme activity^{3,7}. However, in these studies only the NADP⁺-dependent oxidative decarboxylation of isocitrate to α -ketoglutarate was assessed. To understand how IDH1 activity is altered in cells by the presence of R132H mutant IDH1, we expressed increasing

amounts of wild-type and R132H mutant IDH1 separately or in combination and assessed isocitrate-dependent NADPH production and α -ketoglutarate-dependent NADPH consumption in cell lysates. Consistent with published results, expression of R132H mutant IDH1 resulted in no measurable production of NADPH from isocitrate, and isocitrate-dependent NADPH production increased with increasing amounts of wild-type enzyme (Supplementary Fig. 3a, b). The ability of the wild-type enzyme to generate NADPH was decreased slightly by co-expression of the R132H mutant IDH1. Opposite results were obtained, however, when NADPH consumption was measured in the presence of α-ketoglutarate. NADPH consumption by wild-type enzyme was not observed, whereas R132H mutant IDH1 expression resulted in α-ketoglutarate-dependent NADPH consumption (Supplementary Fig. 3c). Although the overall consumption of NADPH was slow, if anything co-expression of wildtype IDH1 with R132H mutant IDH1 facilitated the α-ketoglutaratedependent consumption of NADPH. These findings demonstrate that in contrast to wild-type IDH1, R132H mutant IDH1 promotes an NADPH-dependent reduction of α-ketoglutarate. Furthermore, as this reduction was not inhibited by co-expression of wild-type IDH1, these data indicate that the novel activity of mutant IDH1 can persist even in the presence of a wild-type *IDH1* allele. In fact, it is possible that in the case of a heterodimer of wild-type and mutant IDH1, the α-ketoglutarate and NADPH produced locally by the wildtype subunit could be used as substrates for the mutant subunit, explaining the decrease in NADPH production observed in lysates when wild-type and mutant IDH1 are co-expressed.

To understand how R132 mutations alter the enzymatic properties of IDH1, wild-type and R132H mutant IDH1 proteins were produced and purified from *Escherichia coli*. When NADP⁺-dependent oxidative decarboxylation of isocitrate was measured using purified wild-type or R132H mutant IDH1 protein, it was confirmed that R132H mutation impairs the ability of IDH1 to catalyse this reaction^{3,7}, as evident by the loss in binding affinity for both isocitrate and

Table 1 R132H mutation alters the enzymatic properties of IDH1

Kinetic parameter of reaction	Wild-type IDH1	R132H IDH1
Oxidative (→NADPH)		
$K_{\text{m.NADP}}^+$ (μ M)	49	84
K _{m.isocitrate} (μΜ)	65	370
$K_{m,MgCl2}$ (μ M)	29	1.0×10^{4}
$K_{i,\alpha\text{-ketoglutarate}}$ (μ M)	1.9×10^{3}	24
$k_{\rm cat}$ (s ⁻¹)	4.4×10^{4}	37.5
Reductive (\rightarrow NADP ⁺)		
$K_{m,NADPH}$ (μM)	n/a*	0.44
$K_{m,\alpha\text{-ketoglutarate}}$ (μ M)	n/a	965
$k_{\text{cat}} (s^{-1})$	n/a	1.0×10^{3}

Kinetic parameters of oxidative and reductive reactions as measured for wild-type and R132H IDH1 enzymes are shown. K_m (Michaelis constant) and k_{cat} values for the reductive activity of the wild-type enzyme were unable to be determined as no measurable enzyme activity was detectable at any substrate concentration. K_n inhibition constant.

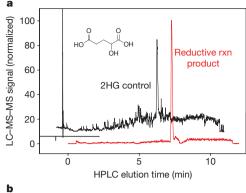
MgCl₂ along with a 1,000-fold decrease in catalytic turnover (Table 1 and Supplementary Fig. 4a). In contrast, when NADPH-dependent reduction of α -ketoglutarate was assessed using either wild-type or R132H mutant IDH1 protein, only R132H mutant could catalyse this reaction (Table 1 and Supplementary Fig. 4b). Part of this increased rate of α -ketoglutarate reduction results from an apparent increase in affinity for both the cofactor NADPH and substrate α -ketoglutarate in the R132H mutant IDH1 (Table 1). Taken together, these data demonstrate that whereas the R132H mutation leads to a loss of enzymatic function for oxidative decarboxylation of isocitrate, this mutation also results in a gain of enzyme function for the NADPH-dependent reduction of α -ketoglutarate.

Reduction of the α -ketone in α -ketoglutarate can result in 2HG. To determine whether R132H mutant protein directly produced 2HG from α-ketoglutarate we examined the product of the mutant IDH1 reaction using negative ion mode triple-quadrupole electrospray LC-MS. These experiments confirmed that 2HG was the direct product of NADPH-dependent α-ketoglutarate reduction by the purified R132H mutant protein through comparison with known metabolite standards (Fig. 2a). Conversion of α -ketoglutarate to isocitrate was not observed. To determine the chirality of the 2HG produced, we derivatized the products of the R132H reaction with diacetyl-L-tartaric anhydride, which allowed us to separate the (S) and (R) enantiomers of 2HG by simple reverse-phase LC and detect the products by tandem mass spectrometry9 (Fig. 2b). The peaks corresponding to the (S) and (R) isomers of 2HG were confirmed using racemic and R(-)-2HG standards. The reaction product from R132H co-eluted with the R(-)-2HG peak, demonstrating that the R(-) stereoisomer is the product produced from α -ketoglutarate by R132H mutant IDH1.

To determine whether the altered enzyme properties resulting from the R132H mutation were shared by other R132 mutations found in human gliomas, recombinant R132C, R132L and R132S mutant IDH1 proteins were generated and the enzymatic properties assessed. Similar to R132H mutant protein, R132C, R132L and R132S mutations all result in a gain-of-function for NADPH-dependent reduction of α -ketoglutarate (Supplementary Fig. 4). Thus, in addition to impaired oxidative decarboxylation of isocitrate, one common feature shared among the IDH1 mutations found in human gliomas is the ability to catalyse direct NADPH-dependent reduction of α -ketoglutarate.

X-ray structure reveals a distinct active site of mutant IDH1

To define how R132 mutations alter the enzymatic properties of IDH1, the crystal structure of R132H mutant IDH1 bound to α -ketoglutarate, NADPH and Ca^{2+} was solved at 2.1 Å resolution (see Supplementary Table 1 for crystallographic data and refinement statistics). The overall quaternary structure of the homodimeric R132H mutant enzyme adopts the same closed catalytically competent conformation (shown as a monomer in Fig. 3a) that has been previously described for the wild-type enzyme¹⁰.



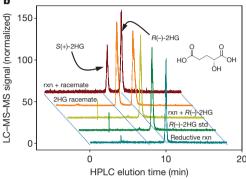


Figure 2 | **R132H** mutation in IDH1 results in production of R(-)-2HG. **a**, 2HG was identified as the reductive reaction product of recombinant human R132H mutant IDH1 using LC–MS as shown. **b**, The chirality of 2HG produced by R132H mutant IDH1 was assessed as by diacetyl-L-tartaric anhydride derivatization and LC–MS analysis. Normalized LC–MS signal for the reductive reaction (rxn) product alone, an R(-)-2HG standard alone, and the two together (rxn + R(-)-2HG) are shown, as is the signal for a racemic mixture of R(-) and S(+) forms (2HG racemate) alone or with the reaction products (rxn + racemate).

Two important features were noted by the change of R132 to histidine: the effect on conformation equilibrium and the reorganization of the active site. Located atop a β -sheet in the relatively rigid small domain, R132 acts as a gate-keeper residue and seems to orchestrate the hinge movement between the open and closed conformations. The guanidinium moiety of R132 swings from the open to the closed conformation with a distance of nearly 8 Å. Substitution of histidine for arginine is likely to change the equilibrium in favour of the closed conformation that forms the catalytic cleft for cofactor and substrate to bind efficiently, which partly explains the high affinity for NADPH shown by the R132H mutant enzyme. This feature may be advantageous for the NADPH-dependent reduction of α -ketoglutarate to R(-)-2HG in an environment where NADPH concentrations are low. Second, closer examination of the catalytic pocket of the mutant IDH1 structure in comparison to the wild-type enzyme showed not only the expected loss of key salt-bridge interactions between the guanidinium of R132 and the α/β carboxylates of isocitrate, as well as changes in the network that coordinates the metal ion, but also an unexpected reorganization of the active site. Mutation to histidine resulted in a significant shift in position of the highly conserved residues Y139 from the A subunit and K212' from the B subunit (Fig. 3b, c), both of which are thought to be critical for catalysis by this enzyme family¹¹. In particular, the hydroxyl moiety of Y139 now occupies the space of the β -carboxylate of isocitrate.

The electron density in the active site was not sufficient to assign α -ketoglutarate and its orientation unambiguously. We have modelled the substrate based on the available electron density, taking into consideration the coordination between the carbonyl oxygen of α -ketoglutarate and the calcium ion as well as an orientation of α -ketoglutarate that would produce R(-)-2HG, the experimental product. The model required a significant repositioning of α -ketoglutarate compared to isocitrate, such that the distal carboxylate of

^{*} No measurable enzymatic activity.

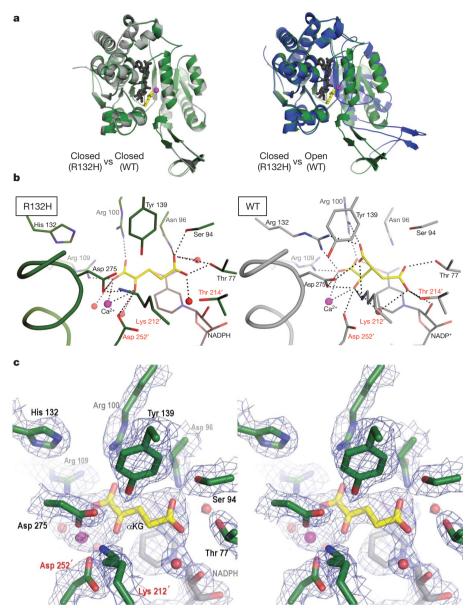


Figure 3 | Structural analysis of R132H mutant IDH1. a, On the left is shown an overlay of R132H mutant IDH1 (green) and wild-type IDH1 (grey) structures in the 'closed' conformation. On the right is shown an overlay of wild-type IDH1 (blue) structure in the 'open' conformation with mutant IDH1 (green) for comparison. b, Close-up comparison of the R132H IDH1 active site (left) with α -ketoglutarate (yellow) and NADPH (grey) and the wild-type IDH1 active site (right) with isocitrate (yellow) and NADP (grey).

Residues coming from the other monomer are denoted with a prime (') symbol. In addition to the mutation at residue 132, the major changes are the positions of the catalytic residues Tyr 139 and Lys 212'. **c**, Wall-eyed stereo image showing the composite omit map for α -ketoglutarate, NADPH, calcium ion, His 132 and other key catalytic residues in the R132H mutant active site contoured at the 1σ level.

 α -ketoglutarate now points upward to make new contacts with N96 and S94. Overall, this single R132H mutation results in formation of a distinct active site compared to wild-type IDH1.

2HG levels are elevated in human glioma samples

Our data demonstrate that mutation of R132 can result in the ability of IDH1 to generate R(-)-2HG from α -ketoglutarate. To determine if 2HG production is characteristic of tumours harbouring mutations in IDH1, metabolites were extracted from human malignant gliomas that were either wild-type or mutant for IDH1 (see Supplementary Table 2 for summary of tumour characteristics). It has been suggested that α -ketoglutarate levels are decreased in cells transfected with mutant IDH1 7 . We observe that the average α -ketoglutarate level from 12 tumour samples harbouring various R132 mutations was slightly less than the average α -ketoglutarate level observed in 10 tumours which are wild type for IDH1. This difference

in α-ketoglutarate was not statistically significant, and a range of α-ketoglutarate levels was observed in both wild-type and mutant tumours (Fig. 4). Similarly, a range of levels was observed for other proximal TCA cycle metabolites with no significant differences observed between wild-type IDH1 tumours and tumours with R132 IDH1 mutations. In contrast, increased 2HG levels were found in all tumours that contained an R132 IDH1 mutation (Fig. 4 and Supplementary Table 2). All R132 mutant IDH1 tumours examined had between 5 and 35 μmol of 2HG per gram of tumour, whereas tumours with wild-type IDH1 had over 100-fold less 2HG. This increase in 2HG in R132 mutant tumours was statistically significant (P < 0.0001). We confirmed that (R)-2HG was the isomer present in tumour samples (data not shown). Together these data establish that the novel enzymatic activity associated with R132 mutations in IDH1 results in the production of 2HG in human brain tumours that harbour these mutations.

ARTICLES

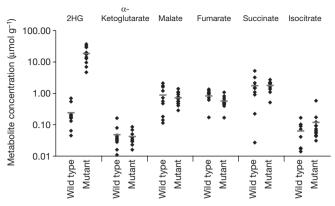


Figure 4 | Human malignant gliomas containing R132 mutations in IDH1 contain increased concentrations of 2HG. Human glioma samples obtained by surgical resection were snap frozen, genotyped to stratify as wild type (n=10) or carrying an R132 mutant allele (mutant) (n=12) and metabolites extracted for LC–MS analysis. Among the 12 mutant tumours, 10 carried a R132H mutation, one an R132S mutation, and one an R132G mutation. Each symbol represents the amount of the listed metabolite found in each tumour sample. Horizontal lines indicate the group sample means. The difference in 2HG observed between wild-type and R132 mutant IDH1 tumours was statistically significant by Student's t-test (P < 0.0001). There were no statistically significant differences in α-ketoglutarate, malate, fumarate, succinate, or isocitrate levels between the wild-type and R132 mutant IDH1 tumours.

Discussion

2HG is known to accumulate in the inherited metabolic disorder 2-hydroxyglutaric aciduria. This disease is caused by deficiency in the enzyme 2-hydroxyglutarate dehydrogenase, which converts 2HG to α-ketoglutarate¹². Patients with 2-hydroxyglutarate dehydrogenase deficiencies accumulate 2HG in the brain as assessed by MRI and CSF analysis, develop leukoencephalopathy, and have an increased risk of developing brain tumours^{13–15}. Furthermore, elevated brain levels of 2HG result in increased ROS levels^{16,17}, potentially contributing to an increased risk of cancer, and alterations in NADPH metabolism resulting from mutant IDH1 expression could further exacerbate this effect. The ability of 2HG to act as an NMDA (Nmethyl-D-aspartate) receptor agonist may contribute to this effect¹⁶. 2HG may also be toxic to cells by competitively inhibiting glutamate and/or α-ketoglutarate using enzymes. These include transaminases, which allow utilization of glutamate nitrogen for amino and nucleic acid biosynthesis, and α-ketoglutarate-dependent prolyl hydroxylases, such as those that regulate Hif1α levels. Alterations in Hif1α have been reported to result from mutant IDH1 protein expression⁷. Regardless of the mechanism, it seems likely that the gain-of-function ability of cells to produce 2HG as a result of R132 mutations in IDH1 contributes to tumorigenesis. Patients with 2-hydroxyglutarate dehydrogenase deficiency have a high risk of central nervous system (CNS) malignancy¹⁵. The ability of mutant IDH1 to act directly on α-ketoglutarate may explain the prevalence of IDH1 mutations in tumours from CNS tissue, which are unique in their high level of glutamate uptake and its ready conversion to α -ketoglutarate in the cytosol¹⁸, thereby providing high levels of substrate for 2HG production. Myeloid cells also display a high ability to metabolize glutamine and recently R132 IDH1 mutations have also been described in a subset of acute myelogenous leukaemia (AML)19. The apparent codominance of the activity of mutant IDH1 with that of the wild-type enzyme is consistent with the genetics of the disease, in which only a single copy of the gene is mutated. As discussed above, the wild-type IDH1 could directly provide NADPH and α-ketoglutarate to the mutant enzyme. These data also demonstrate that mutation of R132 to histidine, serine, cysteine, glycine, or leucine shares a common ability to catalyse the NADPH-dependent conversion of α-ketoglutarate to 2HG. These findings help clarify why mutations at other amino acid residues of IDH1, including other residues essential for

catalytic activity, are not found. Finally, these findings have clinical implications in that they suggest that 2HG production will identify patients with IDH1 mutant brain tumours. This will be important for prognosis as patients with IDH1 mutations live longer than patients with gliomas characterized by other mutations⁵. In addition, patients with lower-grade gliomas may benefit by the therapeutic inhibition of 2HG production. Inhibition of 2HG production by mutant IDH1 might slow or halt conversion of lower-grade glioma into lethal secondary glioblastoma, changing the course of the disease.

METHODS SUMMARY

R132H, R132C, R132L and R132S mutations were introduced into human IDH1 by standard molecular biology techniques. 293T and human glioma U87MG and LN-18 cell lines were transfected using standard techniques. Protein expression levels were determined by western blot. Metabolites were extracted from cultured cells and from tissue samples using 80% aqueous methanol (-80 °C) as previously reported8. Metabolite levels in samples were determined by negative mode electrospray LC-MS. For untargeted profiling, extract components were resolved using reverse-phase high-performance liquid chromatography (HPLC) and metabolites were detected in ultra-high resolution mode (resolution \sim 100,000) by stand alone orbitrap MS, collecting at one scan per second over an m/z range of 110–1,100. For targeted evaluation of 2HG, α -ketoglutarate and other TCA intermediates, extracts were resolved by reverse-phase HPLC system and metabolites detected by triple-quadrupole mass spectrometry, using multiple-reaction monitoring. Enzymatic activity in cell lysates was assessed by following a change in NADPH fluorescence over time in the presence of isocitrate and NADP, or α-ketoglutarate and NADPH. For enzyme assays using recombinant IDH1 enzyme, proteins were purified from E. coli using Ni affinity and size-exclusion chromatography. Enzymatic activity for recombinant IDH1 protein was assessed by following a change in NADPH absorbance at 340 nm using a stop-flow spectrophotometer. Chirality of 2HG was determined as described previously9. For crystallography studies, purified R132H IDH1 was pre-incubated with NADPH, calcium chloride and α-ketoglutarate. Crystals were obtained at 20 °C by vapour diffusion equilibration using 3 μl drops mixed 2:1 (protein:precipitant) against a well-solution of MES pH 6.5 and PEG 6000. Patient tumour samples were obtained after informed consent as part of a UCLA IRB-approved research protocol, collected by surgical resection, snap frozen in isopentane cooled by liquid nitrogen and stored at -80 °C. The *IDH1* mutation status of each sample was determined as described previously³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 July; accepted 29 October 2009. Published online 22 November 2009.

- Balss, J. et al. Analysis of the IDH1 codon 132 mutation in brain tumors. Acta Neuropathol. 116, 597–602 (2008).
- Watanabe, T., Nobusawa, S., Kleihues, P. & Ohgaki, H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. Am. J. Pathol. 174, 1149–1153 (2009).
- Yan, H. et al. IDH1 and IDH2 mutations in gliomas. N. Engl. J. Med. 360, 765–773 (2009).
- Hartmann, C. et al. Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. Acta Neuropathol. 118, 469–474 (2009).
- Parsons, D. W. et al. An integrated genomic analysis of human glioblastoma multiforme. Science 321, 1807–1812 (2008).
- Bleeker, F. E. et al. IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade gliomas but not in other solid tumors. Hum. Mutat. 30, 7–11 (2009).
- Zhao, S. et al. Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1α. Science 324, 261–265 (2009).
- Lu, W., Kimball, E. & Rabinowitz, J. D. A high-performance liquid chromatographytandem mass spectrometry method for quantitation of nitrogen-containing intracellular metabolites. J. Am. Soc. Mass Spectrom. 17, 37–50 (2006).
- Struys, E. A., Jansen, E. E., Verhoeven, N. M. & Jakobs, C. Measurement of urinary D- and L-2-hydroxyglutarate enantiomers by stable-isotope-dilution liquid chromatography-tandem mass spectrometry after derivatization with diacetyl-Ltartaric anhydride. Clin. Chem. 50, 1391–1395 (2004).
- Xu, X. et al. Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity. J. Biol. Chem. 279, 33946–33957 (2004).
- Aktas, D. F. & Cook, P. F. A lysine-tyrosine pair carries out acid-base chemistry in the metal ion-dependent pyridine dinucleotide-linked β-hydroxyacid oxidative decarboxylases. *Biochemistry* 48, 3565–3577 (2009).
- Struys, E. A. et al. Mutations in the D-2-hydroxyglutarate dehydrogenase gene cause D-2-hydroxyglutaric aciduria. Am. J. Hum. Genet. 76, 358–360 (2005).

- Kölker, S., Mayatepek, E. & Hoffmann, G. F. White matter disease in cerebral organic acid disorders: clinical implications and suggested pathomechanisms. *Neuropediatrics* 33, 225–231 (2002).
- Wajner, M., Latini, A., Wyse, A. T. & Dutra-Filho, C. S. The role of oxidative damage in the neuropathology of organic acidurias: insights from animal studies. *J. Inherit. Metab. Dis.* 27, 427–448 (2004).
- Aghili, M., Zahedi, F. & Rafiee, E. Hydroxyglutaric aciduria and malignant brain tumor: a case report and literature review. J. Neurooncol. 91, 233–236 (2009).
- Kolker, S. et al. NMDA receptor activation and respiratory chain complex V inhibition contribute to neurodegeneration in D-2-hydroxyglutaric aciduria. Eur. J. Neurosci. 16, 21–28 (2002).
- 17. Latini, A. et al. D-2-hydroxyglutaric acid induces oxidative stress in cerebral cortex of young rats. Eur. J. Neurosci. 17, 2017–2022 (2003).
- Tsacopoulos, M. Metabolic signaling between neurons and glial cells: a short review. J. Physiol. (Paris) 96, 283–288 (2002).
- 19. Mardis, E. R. et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N. Engl. J. Med. 361, 1058–1066 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. K. Suto, R. S. Brown and E. Fontano at Xtal BioStructures for performing crystallographic studies, and S. Wang at ChemPartner for assistance with biochemical experiments. We thank G. Petsko for his review of the structure data. We also thank T. Mak, N. Wu, L. Tartaglia, J. Saunders, F. Salituro and D. Schenkein for discussions and/or comments on the manuscript. Asterand, PLC provided some of the glioma specimens and SeqWright Inc. assisted with genomic DNA SNP analysis. J.D.R. is supported by NIH R21 CA128620.

Author Contributions L.D., D.W.W., S.G., B.D.B., M.A.B., E.M.D., V.R.F., H.G.J., S.J., M.C.K., K.M.M., R.M.P., P.S.W., K.E.Y., J.D.R., L.M.L. and S.M.S. contributed extensively to the work presented in this paper. L.C.C., C.B.T., M.G.V.H. and S.M.S. provided support and conceptual advice. L.D., M.G.V.H. and S.M.S. wrote the manuscript.

Author Information R132H mutant IDH1 structure files are deposited in the Protein Data Bank under accession code 3INM. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.M.S. (michael.su@agios.com).

doi:10.1038/nature08617 nature

METHODS

Cloning, expression and purification of IDH1 wild type and mutants in E. coli. The open reading frame (ORF) clone of human isocitrate dehydrogenase 1 (cDNA) (IDH1; NM_005896) was purchased from Invitrogen in pENTR221 and Origene in pCMV6. To transfect cells with wild-type or mutant IDH1, standard molecular biology mutagenesis techniques were used to alter the DNA sequence at base pair 395 of the ORF in pCMV6 to introduce base-pair change from guanine to adenine, which resulted in a change in the amino acid code at position 132 from arginine (wild type) to histidine (mutant; or R132H), and confirmed by standard DNA sequencing methods. For 293T cell transfection, wild-type and R132H mutant IDH1 were subcloned into pCMV-Sport6 with or without a C-terminal Myc-DDK-tag. For stable cell line generation, constructs in pCMV6 were used. For expression in E. coli, the coding region was amplified from pENTR221 by PCR using primers designed to add NDEI and XHO1 restrictions sites at the 5' and 3' ends, respectively. The resultant fragment was cloned into vector pET41a (EMD Biosciences) to enable the E. coli expression of C-terminus His8-tagged protein. Site-directed mutagenesis was performed on the pET41a-ICHD1 plasmid using the QuikChange MultiSite-Directed Mutagenesis Kit (Stratagene) to change G395 to A, resulting in the Arg to His mutation. R132C, R132L and R132S mutants were introduced into pET41a-ICHD1 in an analogous way.

Wild-type and mutant proteins were expressed in and purified from the E. coli Rosetta strain (Invitrogen) as follows. Cells were grown in LB (20 μg ml⁻¹ kanamycin) at 37 °C with shaking until OD₆₀₀ reaches 0.6. The temperature was changed to 18 °C and protein expression was induced by adding IPTG to a final concentration of 1 mM. After 12-16h of IPTG induction, cells were re-suspended in lysis buffer (20 mM Tris, pH 7.4, 0.1% Triton X-100, 500 mM NaCl, 1 mM PMSF, 5 mM β-mercaptoethanol, 10% glycerol) and disrupted by microfluidation. The 20,000g supernatant was loaded on metal chelate affinity resin (MCAC) equilibrated with nickel column buffer A (20 mM Tris, pH 7.4, 500 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol) and washed for 20 column volumes. Elution from the column was effected by a 20-column-volume linear gradient of 10% to 100% nickel column buffer B (20 mM Tris, pH 7.4, 500 mM NaCl, 5 mM β-mercaptoethanol, 500 mM imidazole, 10% glycerol) in nickel column buffer A. Fractions containing the protein of interest were identified by SDS-PAGE, pooled and dialysed twice against a 200-volume excess of gel filtration buffer (50 mM Tris pH 7.5, 200 mM NaCl, 5 mM β-mercaptoethanol, 2 mM MnSO₄, 10% glycerol), then concentrated to 10 ml using Centricon (Millipore) centrifugal concentrators. Purification of active dimers was achieved by applying the concentrated eluent from the MCAC column to a Sephacryl S-200 (GE Life Sciences) column equilibrated with gel filtration buffer and eluting the column with 20-column volumes of the same buffer. Fractions corresponding to the retention time of the dimeric protein were identified by SDS-PAGE and pooled for storage at -80 °C.

Cell lines and cell culture. 293T cells were cultured in DMEM (Dulbecco's modified eagles medium) with 10% fetal bovine serum and were transfected using pCMV-6-based IDH1 constructs in six-well plates with Fugene 6 (Roche) or Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Parental vector pCMV6 (no insert), pCMV6-WT IDH1 or pCMV6-R132H were transfected into human glioblastoma cell lines (U87MG and LN-18 (ATCC; HTB-14 and CRL-2610, respectively)) cultured in DMEM with 10% fetal bovine serum. Approximately 24h after transfection, the cell cultures were transitioned to medium containing G418 sodium salt at concentrations of either 500 $\mu g \, {\rm ml}^{-1}$ (U87MG) or 750 $\mu g \, {\rm ml}^{-1}$ (LN-18) to select stable transfectants. Pooled populations of G418-resistant cells were generated and expression of either wild-type IDH1 or R132 IDH1 was confirmed by standard western blot analysis.

Western blot. For transient transfection experiments in 293T cells, cells were lysed 72 h after transfection with standard RIPA buffer. Lysates were separated by SDS–PAGE, transferred to nitrocellulose and probed with goat-anti-IDHc antibody (Santa Cruz Biotechnology, sc49996) or rabbit-anti-Myc tag antibody (Cell Signaling Technology, number 2278) and then detected with HRP-conjugated donkey anti-goat or HRP-conjugated goat-anti-rabbit antibody (Santa Cruz Biotechnology, sc2004). IDH1 antibody to confirm expression of both wild-type and R132H IDH1 was obtained from Proteintech. The IDH2 mouse monoclonal antibody used was obtained from Abcam.

Metabolite extraction. Forty-eight hours before profiling the appropriate culture medium was changed to identical medium with dialysed fetal calf serum. All cells were grown in 10-cm tissue culture dishes, and the medium was replaced again with identical medium 24 h and 1 h before metabolite extraction. All samples were harvested at a non-confluent density $(1.5-2.0\times10^6 \text{ cells})$. Metabolism was quenched and metabolites extracted by aspiration of media and immediate addition of 3.6 ml 80:20 methanol:water at $-80\,^{\circ}\text{C}$, and transfer

to a dry-ice bed to simultaneously lyse cells and quench metabolism. Cell remnants were scraped from the tissue culture dish and transferred, along with the methanol:water, into a 15-ml conical centrifuge tube. The resulting mixture was centrifuged at 14,000g for 20 min, and the supernatant was moved to a new tube. A 1-ml portion of the supernatant was then dried under nitrogen gas, dissolved in 100 $\,\mu$ l of aqueous LC buffer, spun at 13,000g for 10 min to remove any remaining debris and analysed by LC–MS within 24 h, as described below. At the time of harvest, an equivalently treated plate to each of those harvested was used to determine the total cell count, enabling subsequent normalization of LC–MS signal intensities between culture dishes.

Non-targeted LC–MS. Samples were analysed using high-resolution (\sim 100,000 full width at half maximum resolution) stand alone orbitrap mass spectrometry in full-scan mode (1 scan per second, set to high dynamic range), coupled to liquid chromatography via negative mode electrospray ionization ($-3.0\,\mathrm{kV}$). The scan range was picked to avoid abundant ions like phosphate and sulphate entering the ion trap, and was set as follows: 0–3 min, no scanning; 3–5 min, 85–800 m/z; 5.0–6.7 min, 100–800 m/z; 6.7–9.0 min, 85–800 m/z; 9–16 min, 110–1,000 m/z; 16–24 min, 220–1,000 m/z. For chromatography, a Synergi Hydro-RP, 100 mm \times 2 mm, 2.1 μ m particle size column was used (Phenomonex). Solvent A was 10 mM tributylamine and 15 mM acetic acid in 97% water/3% methanol. Solvent B was methanol. The gradient was (t = 0–2.5 min, 0% B; t = 5–7.5 min, 20% B; t = 13 min, 55% B; t = 15.5–18.5 min, 95% B; t = 19–25 min, 0% B). The flow rate was 200 μ l min $^{-1}$.

Targeted LC-MS, 2HG and TCA metabolite measurements. To detect secreted 2HG in culture media, 500 µl aliquots of conditioned media were collected, mixed 80:20 with methanol, and centrifuged at 3,000 r.p.m. for 20 min at 4 °C. The resulting supernatant was collected and stored at −80 °C before LC–MS to assess 2HG levels. To measure whole-cell associated metabolites, media was aspirated and cells were harvested as described above. Two different liquid chromatography (LC) separation methods were used, each coupled by negative electrospray ionization (ESI, -3.0 kV) to triple-quadrupole mass spectrometers operating in multiple reaction monitoring (MRM) mode, with MS parameters optimized on infused metabolite standard solutions. In both methods, metabolites were separated by reversed-phase chromatography using 10 mM tributylamine as an ion pairing agent in the aqueous mobile phase, according to a variant of a previously reported method²⁰. The first method allowed resolution of TCA metabolites: t = 0,50% B; t = 5,95% B; t = 7,95% B; t = 8,0% B, where B refers to an organic mobile phase of 100% methanol. The second method was specific for 2HG, running a fast linear gradient from 50% to 95% B (buffers as defined above) over 5 min. In both methods, the column was a Synergi Hydro-RP, 100 mm × 2 mm, 2.1 μm particle size (Phenomenex), as above. Metabolites were quantified by comparison of peak areas with pure metabolite standards at known concentration. Metabolite flux studies from ¹³C-glutamine were performed as described previously21.

Detection of isocitrate, α-ketoglutarate and 2HG in purified enzyme reactions by LC–MS. Enzyme reactions performed as described in the text were run to completion as judged by measurement of the oxidation state of NADPH at 340 nm. Reactions were extracted with eight volumes of methanol, and centrifuged to remove precipitated protein. The supernatant was dried under a stream of nitrogen and re-suspended in H_2 O. Analysis was conducted on an API2000 LC–MS–MS (Applied Biosystems). Sample separation and analysis was performed on a 50 × 2 mm, 4 μM Synergi Hydro-RP 80 A column, using a gradient of buffer A (10 mM tributylamine, 15 mM acetic acid, 3% (vol./vol.) methanol, in water) and buffer B (methanol) using MRM transitions.

Cell-lysate-based enzyme assays. 293T cell lysates for measuring enzymatic activity were obtained 48 h after transfection with M-PER lysis buffer supplemented with protease and phosphatase inhibitors. After lysates were sonicated and centrifuged at 12,000g, supernatants were collected and normalized for total protein concentration. To measure IDH oxidative activity, 3 μg of lysate protein was added to 200 µl of an assay solution containing 33 mM Tris-acetate buffer (pH 7.4), 1.3 mM MgCl₂, 0.33 mM EDTA, 100 μM β-NADP, and varying concentrations of D-(+)-threo-isocitrate. Absorbance at 340 nm, reflecting NADPH production, was measured every 20 s for 30 min on a SpectraMax 190 spectrophotometer (Molecular Devices). Data points represent the mean activity of 3 replicates per lysate, averaged among 5 time points centred at every 5 min. To measure IDH reductive activity, 3 µg of lysate protein was added to 200 µl of an assay solution which contained 33 mM Tris-acetate pH7.4, 1.3 mM MgCl₂, 25 μM β-NADPH, 40 mM NaHCO₃, and 0.6 mM α -ketoglutarate. The decrease in 340 nm absorbance over time was measured to assess NADPH consumption, with three replicates per lysate.

Recombinant IDH1 enzyme assays. All reactions were performed in standard enzyme reaction buffer (150 mM NaCl, 20 mM Tris-Cl, pH 7.5, 10 mM MgCl₂ and 0.03% (w/v) bovine serum albumin). For determination of kinetic parameters, sufficient enzyme was added to give a linear reaction for 1 to 5 s. Reaction

doi:10.1038/nature08617 nature

progress was monitored by observation of the reduction state of the cofactor at 340 nm in an SFM-400 stopped-flow spectrophotometer (BioLogic). Enzymatic constants were determined using curve fitting algorithms to standard kinetic models with the Sigmaplot software package (Systat Software).

Determination of chirality of reaction products from enzyme reactions and tumours. Enzyme reactions were run to completion and extracted with methanol as described above, then derivatized with enantiomerically pure tartaric acid before resolution and analysis by LC–MS. After being thoroughly dried, samples were re-suspended in freshly prepared 50 mg ml $^{-1}$ (2*R*,3*R*)-(+)-tartaric acid in dichloromethane:acetic acid (4:1) and incubated for 30 min at 75 °C. After cooling to room temperature, samples were briefly centrifuged at 14,000g, dried under a stream of nitrogen, and re-suspended in H₂O. Analysis was conducted on an API200 LC–MS–MS (Applied Biosystems), using an isocratic flow of 90:10 (2 mM ammonium formate, pH 3.6:MeOH) on a Luna C18(2) 150 × 2 mm, 5 μ M column. Tartaric-acid derivatized 2HG was detected using the 362.9/146.6 MRM transition and the following instrument settings: DP -1, FP -310, EP -4, CE -12, CXP -26. Analysis of the (*R*)-2HG standard, 2HG racemic mixture and methanol-extracted tumour biomass (q.v.) was similarly performed. (*R*)-2HG standard was obtained from Sigma and synthesized internally.

Crystallography. For crystallography studies, purified R132H IDH1 was preincubated with NADPH, calcium chloride and α-ketoglutarate. Crystals were obtained at 20 $^{\circ}$ C by vapour diffusion equilibration of 3 μ l drops of mixed 2:1 (protein:precipitant) against a well solution of MES pH 6.5 and PEG 6000. A diffraction data set with 92% completeness was collected under cryo conditions with 1.08 Å wavelength radiation on an ADSC Quantum 315R detector at Beamline X29A of the NSLS at Brookhaven National Laboratories. The data set was integrated using HKL2000 and scaled using SCALEPACK²². A single molecular replacement solution was obtained with PHASER²³ using published coordinates from wild-type IDH1 in the 'closed' conformation as search model (Protein Data Bank accession code 1T0L). Model building was performed using COOT²⁴ and the final model was refined by the CCP4 program REFMAC5²⁵; graphics were produced using PyMOL²⁶. Water molecules were selectively modelled into the structure based upon strong spherical electron density difference peaks, having proper coordination geometry and distance, located within the first solvent shell layer, and having B-factors <80 and electron density sigma level >1.0. After the final refinement, 92.4% of residues fit in the most favoured regions in the Ramachandran plot, with 6.8% in the additional allowed regions and 0.7% in general allowed regions. The final structure model in the closed form conformation at a resolution of 2.1 Å contains three polypeptide chains of IDH1 R132H, each chain consisting of 425 amino acids (414 residues plus a C-terminal His-tag, with residues 4-410 placed into the model), three NADPH molecules and 301 total waters.

Clinical specimens, metabolite extraction and analysis. The collection of human brain tumour samples was approved by the UCLA IRB. Human brain tumours were obtained during surgical resection, snap frozen in isopentane

cooled by liquid nitrogen, and stored at $-80\,^{\circ}$ C. Clinical classification and grading of the tissue was performed using standard clinical histopathology as established by the WHO. Genomic sequence analysis was deployed to identify brain tumour samples containing either wild-type isocitrate dehydrogenase (IDH1) or mutations altering amino acid 132. Genomic DNA was isolated from 50–100 mg of brain tumour tissue using standard methods. A polymerase chain reaction on the isolated genomic DNA was used to amplify a 295-bp fragment of the genomic DNA that contains both the intron and second exon sequences of human *IDH1* and mutation status assessed by standard molecular biology techniques.

Metabolite extraction was accomplished by adding a 10× volume (m/v ratio) of -80 °C methanol:water mix (80%:20%) to the brain tissue (approximately 100 mg) followed by 30 s homogenization at 4 °C. These chilled, methanolextracted homogenized tissues were then centrifuged at 14,000 r.p.m. for 30 min to sediment the cellular and tissue debris and the cleared tissue supernatants were transferred to a screw-cap freezer vial and stored at $-80\,^{\circ}\text{C}$. For analysis, a 2× volume of tributylamine (10 mM) acetic acid (10 mM) pH 5.5 was added to the samples and analysed by LC-MS as follows. Sample extracts were filtered using a Millex-FG 0.20 µm disk and 10 µl were injected onto a reversephase HPLC column (Synergi 150 mm × 2 mm, Phenomenex) and eluted using a linear gradient LC-MS-grade methanol (50%) with 10 mM tributylamine and 10 mM acetic acid, ramping to 80% methanol:10 mM tributylamine: 10 mM acetic acid over 6 min at 200 µl min⁻¹. Eluted metabolite ions were detected using a triple-quadrupole mass spectrometer, tuned to detect in negative mode with multiple-reaction-monitoring mode transition set according to the molecular masses and fragmentation patterns for eight known central metabolites, including 2-hydroxyglutarate as described above. Data were processed using Analyst Software (Applied Biosystems) and metabolite signal intensities were obtained by standard peak integration methods.

- Luo, B., Groenke, K., Takors, R., Wandrey, C. & Oldiges, M. Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. *J. Chromatogr. A* 1147, 153–164 (2007).
- Munger, J. et al. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nature Biotechnol. 26, 1179–1186 (2008).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode (HKL2000). Methods Enzymol. 276, 307–326 (1997).
- 23. McCoy, A. J. et al. Phaser Crystallographic Software. J. Appl. Cryst. 40, 658–674
- 24. Emsley, P. & Cowtan, K. COOT: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr. D 50, 760–763 (1994).
- 26. DeLano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).

ARTICLES

X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor

Alexander I. Sobolevsky¹, Michael P. Rosconi¹† & Eric Gouaux^{1,2}

lonotropic glutamate receptors mediate most excitatory neurotransmission in the central nervous system and function by opening a transmembrane ion channel upon binding of glutamate. Despite their crucial role in neurobiology, the architecture and atomic structure of an intact ionotropic glutamate receptor are unknown. Here we report the crystal structure of the α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-sensitive, homotetrameric, rat GluA2 receptor at 3.6 Å resolution in complex with a competitive antagonist. The receptor harbours an overall axis of two-fold symmetry with the extracellular domains organized as pairs of local dimers and with the ion channel domain exhibiting four-fold symmetry. A symmetry mismatch between the extracellular and ion channel domains is mediated by two pairs of conformationally distinct subunits, A/C and B/D. Therefore, the stereochemical manner in which the A/C subunits are coupled to the ion channel gate is different from the B/D subunits. Guided by the GluA2 structure and site-directed cysteine mutagenesis, we suggest that GluN1 and GluN2A NMDA (*N*-methyl-D-aspartate) receptors have a similar architecture, with subunits arranged in a 1-2-1-2 pattern. We exploit the GluA2 structure to develop mechanisms of ion channel activation, desensitization and inhibition by non-competitive antagonists and pore blockers.

The development and function of the human brain, and its remarkable capacity for experience-dependent change, hinges on the organization and dynamics of chemical synapses—communication 'contact zones' between neurons. At these specialized sites, chemical transmitters released from presynaptic terminals diffuse across the synaptic cleft and activate receptors localized primarily on the postsynaptic cell¹, thereby transmitting the flow of information from one neuron to another. Glutamate is the predominant chemical transmitter of excitatory synapses in the central nervous system²,³ and receptors for this ubiquitous neurotransmitter are of two classes: metabotropic and ionotropic⁴. Ionotropic glutamate receptors (iGluRs) are ligand-gated ion channels fundamental to neurotransmission at excitatory synapses and are implicated in nearly all aspects of nervous system development and function⁵. iGluRs are also involved in chronic neurodegenerative conditions, in psychiatric disorders and in acute injury or trauma⁶-9.

Comprising the iGluR receptor family are AMPA (GluA1–GluA4), kainate (GluK1–GluK5) and NMDA (GluN1, GluN2A–GluN2D, GluN3A–GluN3B) receptors^{10–14}. Whereas NMDA receptors are obligate heterotetramers¹⁴, AMPA and kainate subunits form functional homotetramers^{10–12,15}, although native receptors are almost exclusively heterotetramers^{16–18}. Each subunit has a modular composition¹⁹ and includes a large extracellular amino-terminal domain (ATD)²⁰ that participates in subtype-specific receptor assembly, trafficking and modulation; a ligand-binding domain (LBD) central to agonist/competitive antagonist binding and to activation gating²¹; a transmembrane domain (TMD) that forms the membrane-spanning ion channel²²; and a cytoplasmic carboxy-terminal domain involved in receptor localization and regulation²³.

AMPA, kainate and NMDA receptors are related in amino acid sequence yet they are divergent with respect to function^{5,24}. Whereas non-NMDA receptors exhibit kinetics of activation, deactivation and desensitization on a millisecond timescale²⁵, NMDA receptors are slower, with corresponding molecular processes occurring on a

timescale of tens to hundreds of milliseconds²⁶. Furthermore, AMPA and kainate receptors only demand glutamate for activation, whereas NMDA receptors function as coincidence detectors, requiring membrane depolarization to relieve magnesium block²⁷ together with binding of glycine²⁸ and glutamate. AMPA receptors sojourn through multiple sub-conductance states contingent upon agonist concentration^{15,29}, indicating independent LBDs and a sequential mechanism of activation³⁰. AMPA and kainate receptors undergo profound desensitization, whereas NMDA receptors desensitize less profoundly and by way of glycine-dependent and glycine-independent mechanisms³¹.

The pharmacology of iGluR family members is distinct. AMPA receptors, for example, are non-competitively antagonized by small molecules binding to the juxta-membrane region³². By contrast, the ATDs of NMDA receptors harbour binding sites for polyamines, protons, zinc ions and ifenprodil^{26,33}. In AMPA receptors, the LBD possesses binding sites for modulators of receptor desensitization and deactivation^{34,35}. All iGluR subtypes, however, possess binding sites for pore blockers within the transmembrane ion channel³⁶.

Despite divergent functional properties, iGluR family members have a common structural design. Clues to the symmetry and architecture of iGluRs derive from studies of isolated domains, demonstrating that ATDs and LBDs assemble as dimeric entities ^{34,37–40}; from electron microscopy on intact receptors, showing an overall two-fold symmetry ^{41,42}; and from amino acid sequence analysis and biophysical studies, indicating a ~4-fold symmetric ion channel ^{19,43}. Absent from our understanding, however, is an accurate, atomic-resolution description of iGluR architecture and symmetry, a definition of subunit arrangement in homomeric AMPA and heteromeric NMDA receptors ^{44,45}, and proof of the symmetry mismatch between the two-fold symmetric extracellular domains and the presumably four-fold symmetric ion channel ³⁴. To answer these fundamental questions, we embarked on crystallographic and functional studies of a full-length eukaryotic iGluR.

Crystallization and structure determination

We exploited fluorescence-detection size-exclusion chromatography (FSEC)⁴⁶ to discover that the rat GluA2 receptor^{10,11}, expressed as the unedited⁴⁷, 'flip' variant⁴⁸, was a promising candidate for structural studies (Supplementary Fig. 3a). We further harnessed FSEC to find that n-undecyl- β -D-thiomaltoside, the competitive antagonist ZK200775 (ref. 49) and a modified receptor polypeptide, termed GluA2_{cryst}, were the optimal detergent, ligand and protein construct for crystallization trials, respectively (Supplementary Figs 1, 2, 3b–d and 4–6). GluA2_{cryst} binds ³H-AMPA with a dissociation constant (K_d) of 3.5 ± 0.5 nM and yields glutamate-gated currents similar to the wild-type receptor (Supplementary Figs 7 and 8). Crystals of GluA2_{cryst} belong to the P1 space group, contain one tetrameric receptor in the unit cell and diffract to 3.6 Å resolution (Supplementary Fig. 9 and Supplementary Table 1).

We solved the GluA2_{cryst} structure by molecular replacement, using the high-resolution structures of the isolated GluA2 ATD³⁹ and the antagonist-bound form of the isolated GluA2 LBD³⁸ as search probes. Phases were improved by multidomain non-crystallographic

symmetry (NCS) averaging, solvent flattening and histogram matching. The resulting electron density maps were of sufficient clarity to build the transmembrane helices comprising the ion channel domain and the linking polypeptides connecting the LBD to the ion channel and the ATD to the LBD (Supplementary Figs 10–12). The weakest density was observed for the S2–M4 linker connecting the LBD and the M4 transmembrane domain. Residues on the cytoplasmic side of the membrane connecting membrane helices M1 to M2, and M2 to M3, were not visible in electron density maps.

We probed the veracity of the polypeptide trace by preparing selenomethionine (SeMet)-labelled receptor of $GluA2_{cryst}$ and of four methionine point mutants at Leu 542 (M1), Gln 586 (Q/R site, M2)⁴⁷, Ile 612 (M3) and Val 800 (M4) and by inspecting the corresponding anomalous difference Fourier maps. We formed crystals with a mercury-labelled variant of ZK200775, measured diffraction data at the Hg L_{III} peak, mapped the antagonist binding site in the LBDs, and found additional mercury binding sites, presumably due to residual methyl mercury chloride (Supplementary Table 1 and Supplementary Figs 13–16). Taken together, the diffraction experiments support the

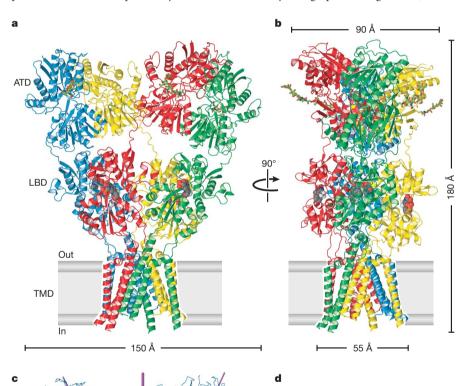


Figure 1 | Architecture of homomeric rat GluA2 receptor. a, View of the 'broad' face of the receptor, perpendicular to the overall two-fold axis of molecular symmetry. Each subunit is in a different colour. b, View of the 'narrow' face of the receptor. In **a** and **b** competitive antagonist molecules lodged within each LBD clamshell are shown in space-filling representation. c, d, Axes of symmetry viewed parallel to the membrane (c) or from the extracellular 'top' of the receptor, along the overall two-fold axis of symmetry (d). Axes of local symmetry for the domains ATD, LBD and TMD are shown in purple, orange and green, respectively. For ATD and LBD, thin lines represent axes of intradimer two-fold symmetry and thick lines represent axes of interdimer twofold symmetry. For TMD, the thick green line represents the local axis of four-fold symmetry. Red mesh peaks (c, d) define mercury sites derived from an anomalous difference Fourier map of a GluA2_{cryst} mercury derivative. The contour level is 5.0σ .

ARTICLES

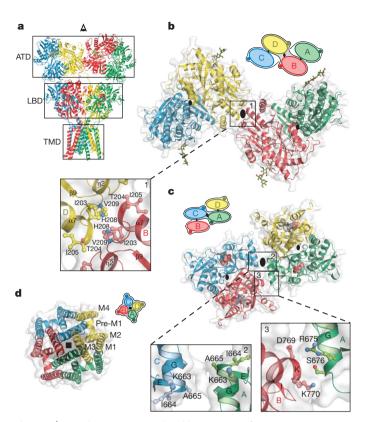


Figure 2 | **Domain symmetry and architecture. a**, $\operatorname{GluA2_{cryst}}$ structure viewed perpendicular to the overall two-fold axis. **b-d**, Domain layers viewed from the top of the receptor, parallel to overall two-fold axis. The simple schematics depict the symmetry and arrangement of domains within each layer. **b**, The ATD layer. The boxed region highlights dimer–dimer contacts, with overall two-fold axis (large black oval) in the centre. The local, intradimer two-fold axes of symmetry are shown as smaller black ovals. Subunits B and D are proximal and subunits A and C are distal to the overall two-fold axis, respectively. **c**, The LBD layer with the dimer–dimer contacts on and off the overall two-fold axis shown in panels 2 and 3, respectively. In this layer, subunits A and C are proximal to the overall two-fold axis. **d**, The TMD layer and its four-fold rotational symmetry (black square).

experimental structure, which was refined to good crystallographic statistics and stereochemistry (Supplementary Table 2).

Architecture and symmetry

The GluA2 receptor has a shape like the capital letter 'Y' where the three major domains are arranged in layers (Fig. 1a, b). The TMDs form the ion channel and define the narrow 'base', the ATDs are splayed outward, like diverging prongs, at the 'top' of the Y, and the LBDs, in complex with antagonist molecules, are sandwiched in between the ion channel and ATDs. In this closed, antagonist-bound state of the receptor there are no prominent vestibules or cavities near the ion channel domain and instead the LBD layer rests like a thick slab on top of the ion channel pore (Supplementary Fig. 17).

The symmetry and subunit arrangement of the tetrameric GluA2 receptor is without precedent. The receptor has an overall yet approximate two-fold axis of molecular symmetry oriented perpendicular to the membrane plane (Fig. 1c, d and Supplementary Fig. 16). This two-fold axis of symmetry relates one ATD dimer³⁹ to another, one LBD dimer³⁴ to the second, and half of the pore-forming TMDs to the other half. The ion channel domain has an approximate four-fold axis of rotational symmetry (Fig. 1c, d).

Each prong of the receptor Y is defined by an ATD dimer in which the 'local' dimer two-fold axis is oriented $\sim\!\!24^\circ$ off of the overall two-fold axis (Fig. 1c, d). 'Below' the ATDs are the LBDs, also organized as a pair of dimers in which the local two-fold axes within each dimer are also tipped off the overall two-fold axis of symmetry by $\sim\!19^\circ$ and are thus not aligned with the 'local' ATD two-folds. 'Below' the LBDs

the four-fold axis of the ion channel is approximately aligned with the overall two-fold axis of symmetry. The multiple, non-aligned axes of local symmetry, together with the two-fold symmetry of the LBDs and the four-fold symmetry of the ion channel, result in symmetry mismatches between the ATDs, LBDs and TMDs.

Extracellular domains

The ATD, implicated in receptor assembly, trafficking and localization, forms two distinct types of subunit–subunit contacts (Fig. 2). On the one hand, within each ATD 'dimer', there are extensive subunit-subunit contacts (A-B or C-D) that are indistinguishable from the contacts seen in the high-resolution crystal structures of the isolated GluA2 ATDs³⁹ (Supplementary Fig. 18a). On the other hand there is an interface between ATD dimers (\sim 330 Å²), located on the overall axis of two-fold symmetry, that is mediated by residues on the L2 lobes of the B and D subunits (Fig. 2b). Within the context of the ATD 'layer', we define the B/D and A/C subunits as 'proximal' and 'distal' to the overall two-fold axis of symmetry. The B–D dimer–dimer interface, although small, is observed in the packing of GluA2 ATD dimers in crystals of the isolated ATD³⁹ (Supplementary Fig. 18b). We suggest that the ATD B-D interface, together with the subunit 'crossover' between the ATD and LBD layers, provides a molecular explanation for the role that the ATD has in the assembly and stability of the tetrameric receptor.

At the LBD layer each agonist binding domain is also a partner in readily identifiable 'dimers' and these dimers, in turn, interact across the overall two-fold axis (Fig. 2c). As a consequence of the subunit crossover between the ATD and LBD layers, the local LBD dimers are formed by the A–D and B–C subunits, with the A and C subunits and the B and D subunits proximal and distal to the overall two-fold axis, respectively. Within a LBD dimer there are multiple contacts between domain 1 of each subunit, recapitulating the interactions seen in the high-resolution crystal structures of the isolated, water-soluble GluA2 LBD³⁸. In the

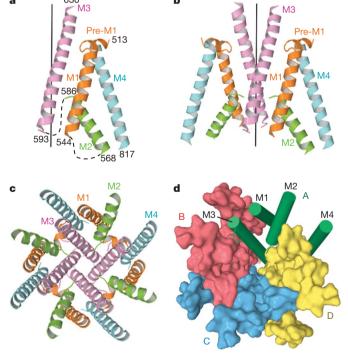


Figure 3 | Transmembrane domain architecture. a, b, Fold of transmembrane domain for subunit A (a) and for subunits A and C (b) viewed parallel to the membrane. Transmembrane segments M1 to M4 are depicted in different colours. The 'vertical' black line defines the fourfold symmetry axis. Dashed lines indicate disordered regions. c, d, TMD viewed parallel to the four-fold axis. c, Transmembrane segments are coloured as in panels a, b. d, Surface representation of subunits B to D. For subunit A, the segments M1–M4 are shown as green cylinders.

GluA2_{cryst} structure, which corresponds to an antagonist-bound, non-desensitized state, the domain 1–domain 1 interface is 'intact', as visualized in the wild type³⁸, the Leu 483 to Tyr mutant or the cyclothiazide³⁴ and aniracetam-bound³⁵ structures of the isolated LBD. Domain 2, by contrast, does not participate in significant intersubunit interactions within a LBD dimer, a finding that is also in harmony with high-resolution studies of the isolated LBD⁵⁰. Lodged in the 'clamshell' of each LBD is a bound antagonist, thus proving that the agonist/competitive antagonist binding site is located within and not between subunits.

There is a small (\sim 224 Å²) interface between LBD dimers, an area consistent with weak dimer–dimer interactions^{34,51}. Like the ATD dimer–dimer interface, the LBD dimer interface is located on the overall two-fold axis of symmetry and is composed primarily of residues at the 'bend' between helices F and G, with the α -carbon atoms of residues Lys 663, Ile 664 and Ala 665 of subunits A and C 8–13 Å apart. The relationship between LBD dimers is not perfectly two-fold symmetric, however, and whereas residues at the end of helices G and K of the A and B subunits (Ser 676 and Lys 770) are in van der Waals contact (α -C α distance of 6.1 Å), the equivalent residues in subunits C and D are not in contact (α -C α distance of 9.7 Å).

Transmembrane domain

Viewed from the extracellular side of the membrane down the overall two-fold axis of symmetry, four GluA2_{cryst} subunits arrange their transmembrane domains around an axis of ~4-fold rotational symmetry (Fig. 2d). In accord with topology studies^{52,53}, each subunit has three transmembrane helices (M1, M3 and M4), a central pore-like helix (M2)⁵⁴ and a polypeptide pore-lining loop that is disordered in our electron density maps (Fig. 3a, b). Leading from the LBD, the S1–M1 polypeptide segment adopts an extended conformation until reaching the TMD, at which point the polypeptide forms a ~90° turn and initiates a short helix (pre-M1) oriented nearly parallel to the membrane. The pre-M1 helix acts like a cuff around the 'top' of the ion channel domain, making contacts with carboxyl and aminoterminal ends of helices M3 and M4, respectively. M1 is the first bona fide transmembrane segment and it resides on the exterior of the ion

channel domain (Fig. 3c). Within the pore lies the M2 helix, positioned largely on the basis of tube-shaped electron density and the anomalous difference density peak from the SeMet-labelled Gln 586 to Met (Q/R site) mutant. The M3 helices line the inside of the ion channel domain, are $\sim\!52\,\text{Å}$ in length and, in the present, antagonist bound structure, cross at the level of the pre-M1 cuff helices, near the membrane—aqueous solution boundary, forming a $\sim\!12\,\text{Å}$ occlusion of the putative ion permeation pathway. Residing on the exterior of the ion channel domain is the M4 helix, connected to the S2 segment of the LBD by two turns of helix and a short extended region of polypeptide. There are extensive subunit—subunit interactions between the transmembrane segments with the M4 segment of one subunit making interactions primarily with the transmembrane domains of an adjacent subunit (Fig. 3d). These interactions provide a molecular basis for the crucial role of the M4 helix in receptor assembly and function 22 .

Probing subunit interfaces

Cognisant that the GluA2_{cryst} structure (Fig. 4a) possesses an unorthodox subunit arrangement and molecular symmetry, we tested whether the subunit arrangement and domain–domain contacts in the crystal structure reflect interactions adopted by the receptor in a noncrystalline environment. To accomplish this, we introduced cysteine residues into the three unexplored interfaces at sites that should result in spontaneous disulphide bond formation (Fig. 4b–d). We purified the mutant, tetrameric receptors to homogeneity and probed the extent of spontaneous subunit crosslinking by gel electrophoresis under reducing and non-reducing conditions (Fig. 4e–g and Supplementary Fig. 19).

At the ATD, dimer–dimer, B to D subunit interface, we introduced a cysteine at Val 209 (Fig. 4b). For the wild-type-like receptor, in the presence or absence of reducing agent, the GluA2 subunit migrates at a position consistent with its calculated molecular mass. By contrast, for the Val 209 to Cys mutant, we observe reducing-agent-dependent dimer formation, thus supporting the presence of this dimer–dimer interface in the intact receptor under native conditions (Fig. 4e). In the LBD layer, there are two distinct subunit–subunit interfaces. One

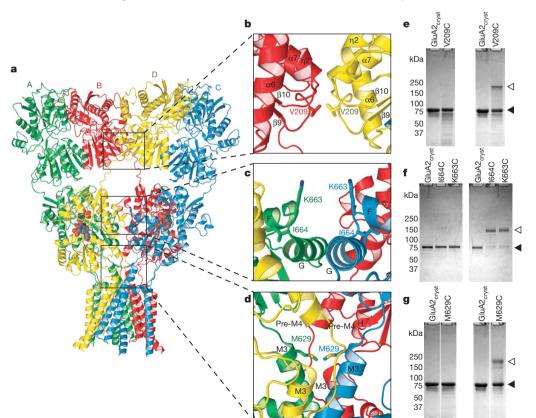


Figure 4 | Probing intersubunit interfaces in GluA2 AMPA receptors. a, Ribbon diagram of the GluA2_{cryst} structure with each subunit in a different colour. **b**–**d**, Close-up views of intersubunit interfaces between two ATD dimers (b), two LBD dimers (c) and at the top of the ion channel (d). e-g, SDS-PAGE analysis of spontaneous crosslinking of cysteines introduced at intersubunit interfaces. Left and right panels illustrate experiments carried out in reducing and nonreducing conditions, respectively. Filled and open triangles indicate positions of monomeric and dimeric bands, respectively.

interface is within a LBD dimer and is formed by extensive contacts between domain 1 of two-fold related subunits, faithfully mirroring the thoroughly documented dimer interface observed in the isolated LBDs of AMPA^{34,38,51}, kainate⁵⁵ and NMDA^{56,57} receptors. The second interface, between subunits (A and C) proximal to the overall two-fold axis, is composed of only a handful of intersubunit contacts (Fig. 4c). We therefore tested whether residues in this interface could form inter-dimer disulphide crosslinks. At both Lys 663 and Ile 664 cysteine mutants formed redox-dependent dimers (Fig. 4f), supporting the presence of this interface in the intact receptor. This LBD dimerdimer interface is also important for agonist-dependent gating because steady-state currents of the Ile 664 to Cys mutant are potentiated ~5-fold after receptor reduction⁵⁸ (Supplementary Fig. 20).

The apex of the ion channel domain, defined by the C-terminal ends of M3, provides an important test of the GluA2_{crvst} structure, not only because residues at the end of M3 define the gate of the ion channel in this antagonist-bound state, but also because the end of M3 and the M3–S2 linker span the transition between the four-fold symmetry of the ion channel and the two-fold symmetry of the extracellular domains. In two subunits (A and C), the Met 629 side chains point towards each other and the α -carbons are separated by $\sim 12 \text{ Å}$, whereas the corresponding α -carbon atoms in subunits B and D are 30 Å apart (Fig. 4d). In satisfying agreement with the GluA2_{cryst} structure, we find that cysteines introduced at position 629 form redox-dependent crosslinks (Fig. 4g), consistent with the proximity of Met 629 residues and the structure of the ion channel domain. We may not see complete crosslinking to a dimer position because of the overall two-fold symmetry of the extracellular domains or because some of the cysteine residues may have suffered chemical modification during expression or purification.

We propose that the architecture of kainate receptors is similar to that of the $GluA2_{cryst}$ AMPA receptor based on the remarkable observation that the isolated GluK2 ATD dimer forms a similar dimer-of-dimers arrangement in the crystal lattice, yielding a 'tetrameric' complex⁴⁰ like that in the $GluA2_{cryst}$ structure (Supplementary Fig. 21). Superposition of α -carbon atoms for 292 residues per subunit in the structurally conserved regions of $GluA2_{cryst}$ and isolated GluK2 ATD dimer-of-dimers yields a root mean square deviation (r.m.s.d.) of 3.2 Å. This observation, in combination with the fact that LBDs of AMPA and kainate receptors form similar local dimers⁵⁵, demonstrates that principles of architecture and symmetry are conserved between AMPA and kainate receptors.

Architecture of NMDA receptors

NMDA receptors are the most complex subfamily of ionotropic glutamate receptors, not only because they are obligate heterotetramers, requiring a glycine-binding GluN1 or GluN3 subunit together with a glutamate binding GluN2 subunit, but also because their ATDs bind ions and molecules that modulate receptor activity²⁶. Even though the structure of the GluN1–GluN2A LBD heterodimer is known⁵⁶, there is no conclusive experimental knowledge of how subunits are arranged in heterotetrameric NMDA receptors. To determine whether the GluA2_{cryst} structure provides a paradigm for understanding NMDA receptor architecture, we carried out cysteine-directed crosslinking experiments on the rat GluN1–GluN2A NMDA receptor.

Guided by superpositions of the GluN1–GluN2A LBD heterodimer⁵⁶ onto the GluA2_{cryst} structure, we designed cysteine substitutions at putative interdimer interfaces. There are three possible arrangements of the GluN1–GluN2A LBD heterodimers within a heterotetrameric receptor and we can distinguish between these models depending on whether redox-dependent crosslinking requires cysteine substitutions in only GluN1, in only GluN2A or in both GluN1 and GluN2A subunits (Supplementary Figs 22 and 23). Remarkably, only a single cysteine substitution in GluN1, at Glu 699, is required to crosslink NMDA receptor subunits at the LBD level (Fig. 5a). Indeed, none of the six cysteine substitutions in GluN2A promoted crosslinking, whereas simultaneous substitutions in both GluN1 and GluN2A resulted in crosslinking similar to

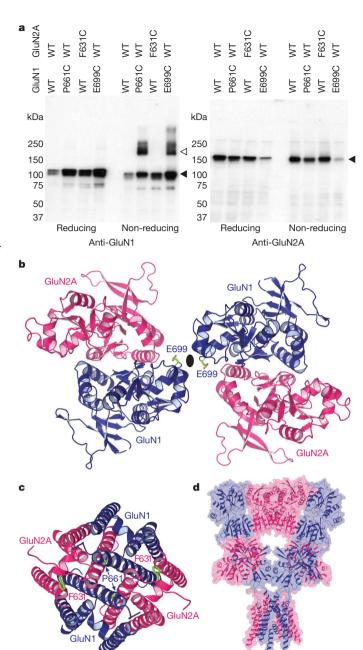


Figure 5 | Subunit arrangement in NMDA receptors. a, Western blot analysis of crosslinking of wild-type GluN1–GluN2A (WT) and cysteine-substituted NMDA receptors probed with anti-GluN1 (left) and anti-GluN2A (right) antibodies. Filled and open triangles indicate positions of uncrosslinked monomeric and crosslinked dimeric subunits, respectively. b, Model of LBD dimer-of-dimers built by superposing two GluN1–GluN2A dimers⁵⁶ on to the GluA2_{cryst} structure, viewed along the axis of overall two-fold symmetry. Residues substituted with cysteines are shown in green. c, Model of the NMDA receptor ion channel based on the GluA2_{cryst} structure and viewed along the axis of pseudo-four-fold symmetry. d, Simple model of NMDA receptor architecture based on the GluA2_{cryst} structure and using the GluN1–GluN2A LBD heterodimer structure together with the GluA2_{cryst} ATDs and TMDs.

substitutions in GluN1 alone. Within the LBD layer of the NMDA receptor, therefore, we propose that GluN1 and GluN2A subunits form diagonal pairs with the GluN1 subunit proximal to the overall two-fold axis of symmetry (Fig. 5b).

What is the arrangement of NMDA receptors within their TMD? Here we exploit the GluN1 and GluN2A equivalents of GluA2 residue Met 629. Because the GluN1 subunit is proximal to the overall two-fold axis of symmetry at the LBD layer, we predict that the GluN1

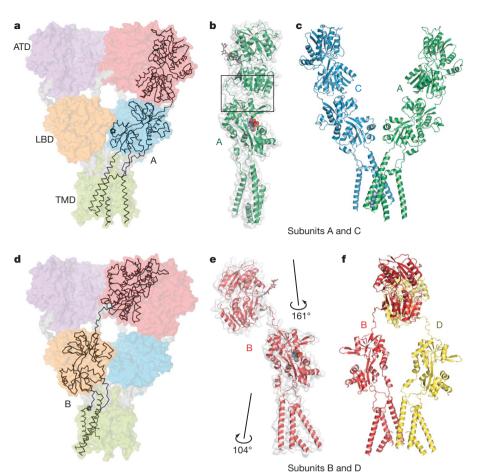


Figure 6 | Subunit non-equivalence and 'domain swapping'. a, α -Carbon trace of subunit A and partially transparent solvent-accessible surface of the entire receptor. b, Conformation of subunit A. c, Position and conformation of subunits A and C in the intact receptor. d, α -Carbon trace of subunit B showing 'domain swapping' in going from the ATD to the LBD layers. e, Conformation of subunit B. f, Position and conformation of subunits B and D in the intact receptor. Note the large differences of the ATD-LBD linkers and ATD-LBD interfaces between the A/C and B/D subunit pairs.

residue, Pro 661, will be proximal to the overall two-fold axis at the TMD layer and that the GluN2A residue, Phe 631, will be too far away to form a crosslink (Fig. 5c). In concordance with this prediction, GluN1(Pro661Cys)—GluN2A(wild type) but not GluN1(wild type)—GluN2A(Phe631Cys) receptors demonstrate redox-dependent crosslinking (Fig. 5a). Even though we have not explored crosslinking of NMDA receptors' ATDs, based on our observation that subunits switch proximity to the overall two-fold axis between the ATD and LBD layers, we predict that the ATDs are assembled as 'local' heterodimers and that NR2 ATDs mediate the dimer—dimer contacts, proximal to an overall two-fold axis of symmetry (Fig. 5d). This arrangement would appropriately position the L2 lobe of the GluN2A ATD to alter receptor structure and function by either modulating ATD dimer—dimer contacts or ATD—LBD interactions, ultimately influencing ion-channel gating via perturbation of the LBD^{26,57}.

Subunit non-equivalence and domain swapping

There is a remarkable swapping of domains, involved in local dimers between subunits, that is illustrated by tracing neighbouring polypeptide chains through the receptor. Within the ATD layer, subunits A and B interact with each other to form a local ATD dimer (red dimer, Fig. 6a, d). When the polypeptide chains of subunits A and B pass to the LBD layer, however, subunit A forms a LBD dimer with the corresponding domain of subunit D (blue dimer, Fig. 6a), while subunit B associates with subunit C (orange dimer, Fig. 6d). Within the TMD, the transmembrane helices of subunits A and B form extensive contacts with each other as well as with subunits C and D (Fig. 3c, d).

The swapping of extensive local dimer interactions between subunits and the symmetry mismatch between the LBDs and TMDs mean that within this homotetrameric receptor, where each subunit is chemically identical, there are two conformationally different subunit pairs related by the overall axis of two-fold molecular symmetry: subunit A is equivalent to C and subunit B is equivalent to D yet the A/C pair is distinct,

in conformation, from the B/D couple (Fig. 6c, f). These differences in conformation can be illustrated by superimposing the LBDs of subunits A and B (Fig. 6b, e), showing the large differences in orientation of the flanking ATDs and TMDs. In subunit A or C, for example, there is a substantial interface between the ATD and LBD (\sim 315 Ų), whereas in subunit B or D there are no similar contacts (Fig. 6b, e). Another fundamental difference is that the LBD of the B/D subunits is \sim 6 Å closer to the putative membrane plane than the A/C subunit pair.

The conformational difference between the two types of subunits is defined by polypeptides linking the ATDs, LBDs and TMDs that have the same amino acid sequence but adopt different conformations. These polypeptides also mediate symmetry transitions between domain layers. The symmetry mismatch between the ATDs and the LBDs, characterized by different orientations of their local two-fold axes of symmetry (Fig. 1c, d) and domain swapping (Fig. 6a, d), is mediated by the ATD–S1 linkers (Thr 377–Lys 393). On the one hand, the ATDs of the A and C subunits interact with their LBDs via an ATD–LBD interface (Fig. 6a–c) and the ATD–S1 linkers adopt a more compact conformation. On the other hand, the ATDs of the B and D subunits are suspended between LBD dimers and the ATD–S1 linker nearly spans a LBD dimer, taking on an extended conformation (Fig. 6d–f).

Two-fold to four-fold symmetry transition

On the extracellular side of the membrane, both the ATD dimers and the LBD dimers are arranged with an overall two-fold symmetry. By contrast, the GluA2_{cryst} TMD has four-fold rotational symmetry (Fig. 3c). Where is the two-fold to four-fold symmetry mismatch structurally reconciled? We can answer this question by examining selenium and mercury positions derived from SeMet-labelled receptor and from crystals grown with the mercury-containing antagonist. Selenium sites at Met 629, a residue within the highly conserved 'Glu-Arg-Met-Val' sequence at the end of M3 immediately preceding the S2 segment, clearly show two-fold symmetry (Fig. 7a, b). Mercury

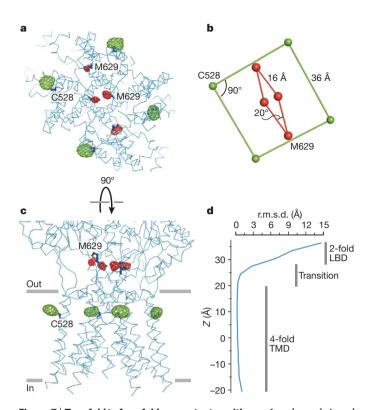


Figure 7 | Two-fold to four-fold symmetry transition. a, Ion channel viewed from the cytoplasm, parallel to the local four-fold axis of symmetry. Mercury (green mesh) and selenium (red mesh) sites, defined by anomalous difference electron density, contoured at 5.0σ and 2.6σ , respectively. b, Geometry of the mercury and selenium sites from panel a. c, View of the TMD perpendicular to its local four-fold axis of molecular symmetry. d, Graph showing the r.m.s.d. values of α -carbon positions following transformation of the A subunit on the B subunit by applying the local four-fold axis of symmetry, emphasizing that the two-fold to four-fold symmetry transition occurs between the LBDs and the membrane-embedded TMD. The 'Z' axis is along the four-fold TMD axis with an origin near the ion channel Gln 586 Q/R site.

sites at Cys 528, near the extracellular end of M1, however, demonstrate unambiguous four-fold symmetry (Fig. 7a, b). By viewing the receptor perpendicular to the overall two-fold axis, and defining the approximate membrane boundary based on residue polarity, we see that Met 629 is just 'outside' of the membrane-spanning region whereas Cys 528 resides within the membrane bilayer. The two-fold to four-fold symmetry transition occurs within this region, a \sim 10-Å thick slab between the LBDs and TMDs (Fig. 7c).

To define precisely where the transition occurs, we applied the four-fold rotational transformation associated with the TMDs to one of the receptor subunits and calculated the r.m.s.d. of $\alpha\text{-}carbon$ atom positions between the superimposed subunits. Where the receptor conforms to four-fold symmetry, the r.m.s.d. values in atom positions are on the order of the coordinate uncertainty. In two-fold symmetric regions, however, the r.m.s.d. values in atom positions are much greater. Thus, by plotting the r.m.s.d. in atom positions as a function of distance along the overall two-fold axis, we can follow the two-fold to four-fold symmetry transition (Fig. 7d). This analysis demonstrates that the transition is abrupt and is located at the boundary of the extracellular leaflet of the membrane bilayer, adjacent to the region of the ion channel domain encircled by the pre-M1 cuff helical segments.

Gating machinery and symmetry mismatch

The symmetry mismatch between the LBDs and the TMDs is resolved by three linking peptides—the S1–M1 (Lys 506–Gly 513), M3–S2 (Val 626–Glu 634) and S2–M4 (Gly 774–Ser 788) linkers—making the transition from the two-fold, parallelogram-like symmetry of

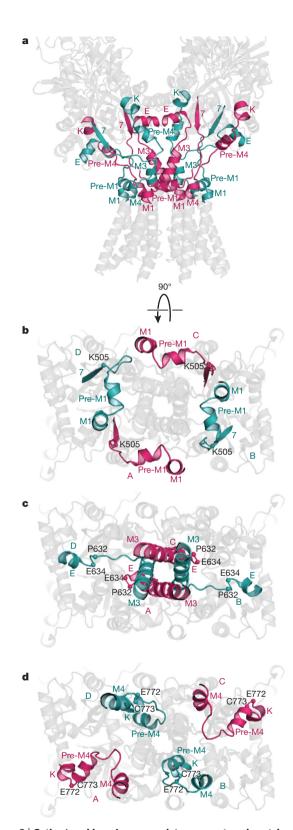


Figure 8 | **Gating 'machinery' accommodates symmetry mismatch. a**, LBDs and TMDs of GluA2_{cryst} viewed perpendicular to the overall two-fold axis of molecular symmetry. The elements mediating symmetry mismatch between LBDs and TMDs—the S1–M1, M3–S2 and S2–M4 linkers—are coloured pink (subunits A and C) or blue (subunits B or D). **b–d**, The elements resolving symmetry mismatch between the LBDs and the TMDs viewed from the cytoplasm, parallel to the ion channel four-fold axis of symmetry. The S1–M1 linkers (**b**), the M3–S2 linkers (**c**) and the S2–M4 linkers (**d**) are shown.

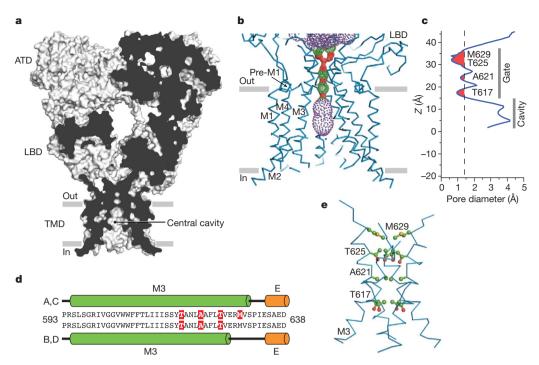


Figure 9 | **Closed conformation of the ion channel pore. a**, Sagittal section of the GluA2_{cryst} receptor illustrates that the occlusion of the putative ion permeation pathway, or the ion channel gate, is located near the extracellular side of the membrane, at the crossing of the M3 helices. **b**, **c**, Surface representation of the ion conduction pathway (**b**) and the pore diameter as a function of distance along the central axis of the channel (**c**) generated using the program HOLE (red < 1.4 Å < green < 2.8 Å < purple). Residues

the LBDs to the four-fold symmetric, square geometry of the TMDs (Fig. 8a). Indeed, these are the central elements of the iGluR gating machinery that transform ligand-induced structural changes in the LBD dimers into the movement of the transmembrane domains that opens and closes the central pore of the four-fold symmetric ion channel. In bridging two-fold to four-fold symmetry transition, the peptide segments linking the LBDs and TMDs can be grouped into two pairs that belong to diagonal subunits A/C or B/D; within each group, the linkers adopt approximately similar conformations whereas between the groups the conformations are clearly distinct.

Accommodation of the two-fold to four-fold symmetry mismatch is illustrated by the different conformations of the S1-M1 linkers which, when passing from the β7 strand of the LBD to M1 (TMD), either come from inside (A/C) or outside (B/D) of the 'M1 circle' (Fig. 8b). The M3–S2 linker provides a particularly striking example of how the conformations of the A/C and B/D linking peptides differ. In the A and C subunits, the M3-S2 linker adopts a helical conformation to Met 629 (Fig. 4d). For the B and D subunits, by contrast, the helical conformation is broken at Val 626 and following this residue, the peptide adopts an extended conformation. The difference in mainchain conformation means that the α -carbon atoms of Pro 632 in the A and C subunits are \sim 27 Å apart whereas the corresponding atoms in the B and D subunits are separated by \sim 50 Å (Fig. 8c). A third striking illustration is provided by Cys 773, a conserved residue at the end of helix K of the LBD. For Cys 773, in subunits B and D, the α -carbons are 33 Å apart and in the A and C subunits, the span is much larger, 69 Å. To reach the M4 in TMD, which possesses four-fold symmetry and where the distances between the same atom in the A/C and B/D subunits are necessarily the same, the A/C and B/D S2-M4 linkers take on different conformations and different orientations relative to the overall two-fold axis (Fig. 8d).

The ion channel

The four-fold rotationally symmetric GluA2_{cryst} ion channel is shaped like a Mayan temple with a broad cytoplasmic base, \sim 42 Å on a side, and

forming the narrowest portions of the ion conduction pathway are indicated. **d**, The A/C and B/D M3 segments adopt distinct conformations at their C termini, proximal to the LBDs. Helical regions are shown as cylinders and coloured green for TMD and orange for LBD. Residues forming the narrowest portions of the ion conduction pathway are highlighted in red. **e**, M3 residues forming the narrowest portion of the ion conduction pathway.

with a bluntly pointed extracellular 'top'. In the present, competitive antagonist-bound state, the ion channel unambiguously adopts a closed conformation by the crossing of the M3 helices (Fig. 9). The crossing of the helices occurs near the highly conserved Ser-Tyr-Thr-Ala-Asn-Leu-Ala-Ala-Phe (SYTANLAAF) motif, with Thr 617, Ala 621 and Thr 625 occluding the ion channel permeation pathway^{22,59}. In the Lurcher mouse⁶⁰ there is a substitution of Ala 636 by Thr in the GluD2 subunit, leading to spontaneously opening ion channels⁶¹. The equivalent residue in GluA2 (Ala 622) participates in close contacts with the M3 helix of a neighbouring subunit, indicating that introduction of bulky residues can directly destabilize the tight helix crossing associated with the resting, closed state of the receptor, leading to constitutively open ion channels. The ion channel permeation pathway is also occluded above the SYTANLAAF motif by a pair of Met 629 residues on the A/C

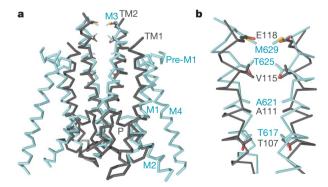


Figure 10 | **Closed state conformations of iGluR and K**⁺ **channels are similar. a, b,** Two diagonal subunits (**a**) and their bundle crossing (**b**) in GluA2_{cryst} channel (blue) and KcsA channel⁵⁴ (grey) viewed parallel to the membrane. Channels were superimposed by aligning main-chain atoms of the M3 and M1 segments in iGluR with the inner and outer helices in KcsA, respectively. Residues forming the narrowest portions of the ion conduction pathway are shown as stick models.

subunits protruding their side chains towards the centre of the pore. Notably, these methionines are adjacent to the Glu 627–Arg 628 motif that, when mutated, strongly perturbs receptor desensitization⁶².

Stimulated by hypotheses¹⁹ and experimental data^{43,63} proposing a common architecture for the ion channel pores of iGluRs and K channels, we superimposed the transmembrane domains of GluA2_{cryst} onto the bacterial K⁺ channel KcsA⁵⁴ (Fig. 10 and Supplementary Fig. 24). Despite the low pairwise identity of \sim 20% between the aligned amino acid sequences of the rat GluA2 receptor channel and KcsA (Supplementary Fig. 25), the M1, M2 and M3 segments of GluA2_{cryst} overlap remarkably well with structurally equivalent portions of KcsA (Fig. 10a). In fact, superposition of α -carbon atoms for 64 residues per subunit in GluA2_{cryst} and KcsA yields a r.m.s.d. of 2.2 Å. Not included in this comparison was transmembrane segment M4, which is absent in KcsA and bacterial glutamate receptors⁶⁴. In addition to the overall similarity, the structural alignment demonstrates that the occlusion of the ion conductive pathway, that is, the regions of these channels involved in gating and defined by the carboxyl terminal region of M3 in GluA2_{crvst} and TM2 in KcsA, is remarkably similar (Fig. 10b). By contrast, the region of KcsA that confers ion selectivity—the extended Thr-Val-Gly-Tyr-Gly selectivity filter—is completely different in amino acid sequence in iGluRs and is disordered in the GluA2_{crvst} structure, observations consistent with the fact that K⁺ channels are highly selective and iGluRs are not.

The pre-M1 cuff helix is oriented nearly parallel to the membrane, at the interface between the membrane and extracellular solution, and is a structural feature that underscores similarities and differences between iGluRs and K⁺ channels⁶⁵. Appearing as a 'collar' around the TMD bundle, the pre-M1 helices may restrain the mobility of M3 in the closed state and, by virtue of its connection via the S1–M1 linker to the LBD, promote opening of the ion channel gate upon agonist binding to the LBD. Whereas some K⁺ channels have a similar element of structure⁶⁵, the length of the helical segment as well as its conformation relative to the ion channel diverge from the GluA2_{cryst} pre-M1 helix. In

fact, Pro 520 in the 'elbow' of the GluA2_{cryst} pre-M1 is highly conserved across iGluR subtypes yet is missing in K⁺ channels (Supplementary Fig. 25), possibly indicating a distinct role of the pre-M1 element in iGluR structure and gating.

Immediately above the Q/R site, clearly within the transmembrane portion of the ion channel domain, is a central cavity (Fig. 9a) similar to that observed in potassium channels⁵⁴. In the GluA2_{cryst} structure there are conspicuous 'gaps' between the transmembrane domains that result in a portal between the central cavity, within the ion channel, and the membrane environment. We speculate that this portal may be occupied by residues projecting from the transmembrane domains of AMPA receptor auxiliary subunits^{66,67}, thus providing a mechanism by which auxiliary subunits modulate ion channel properties, such as the extent of block by polyamines⁶⁸.

The amino acid sequence and structural relationships between GluA2_{cryst} and K⁺ ion channel pores allow us to speculate on the conformation of an iGluR in an open channel state (Supplementary Figs 25 and 26). On the basis of superimposing the ion channels of GluA2_{cryst} and Shaker⁶⁵, we suggest that the transmembrane helices of iGluR will bend and splay away from the central axis of the channel, mimicking the iris-like opening of K⁺ channels. Although roughly similar in overall nature, the TMD movements during activation gating of iGluRs and K+ channels may be different in molecular detail⁶⁹. In fact, the residues lining the ion permeation pathway or the amino acids that perturb gating via genetic or chemical modification, together with the types of non-competitive antagonists and channel blockers that bind to the TMD, both in AMPA (Supplementary Fig. 27) and NMDA (Supplementary Fig. 28) receptors, are highly specific to iGluRs, thus providing further evidence of distinction from potassium channels.

Mechanism of activation

The GluA2_{cryst} structure allows us to interpret decades of studies in the context of an intact receptor. Most fundamentally, the GluA2_{cryst}

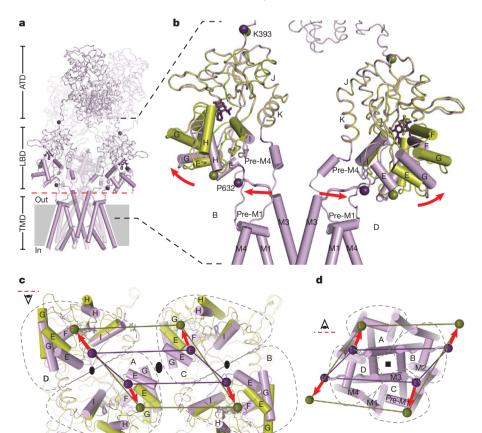


Figure 11 | iGluR activation gating. a, The structure of GluA2_{cryst} with two subunits (A and C) semi-transparent. The red dashed line indicates the interface between LBD and TMD. b, Close-up view of the LBD-TMD regions of subunits B and D. The structure of the watersoluble GluA2 LBD (S1S2) crystallized in complex with glutamate38 has been superimposed, using the D1 domain, on the corresponding region of GluA2_{cryst} and is shown in green. Helical regions of the ion channel as well as parts of the LBD that move upon iGluR activation are shown as cylinders. Purple and green spheres indicate positions of the α -carbons for the residues Lys 393 and Pro 632. Stick models of ZK200775 and glutamate are shown in purple and green, respectively. Red arrows indicate movement during iGluR activation. c, d, Views of the iGluR tetramer from the interface between LBD and TMD (red dashed line in a) on to the LBD along the overall axis of the two-fold symmetry (c) and ion channel along the axis of four-fold symmetry (d). Grey dashed lines outline borders of the A, B, C and D subunits. Purple and green lines connect Pro 632 α-carbon atoms.

structure proves that agonist binding sites are located within individual subunit 'clamshells' and not between subunits, as is the case with trimeric P2X⁷⁰ and pentameric Cys-loop⁷¹ receptors. This architectural principle is consistent with the observations that binding of multiple agonist molecules and subsequent ion channel activation are largely stepwise, sequential processes¹⁵. Indeed, the independence of LBD function is further supported by the fact that they can be genetically excised and studied as isolated soluble domains⁷². Together, these properties justify exploiting the wealth of structural and biophysical experiments on soluble LBDs of AMPA^{38,58,73}, kainate^{55,74} and NMDA receptors^{56,75,76} to illuminate principles of gating in full-length iGluRs.

The GluA2_{crvst} structure is a complex with the high-affinity competitive antagonist ZK200775. Antagonist binding stabilizes the binding domain clamshell in an open conformation (Supplementary Fig. 29)38 and, in the context of the LBD dimer, places the transmembrane-associated linker regions closest together (Supplementary Fig. 30). The conformation of the LBD trapped by ZK200775, although similar to the apo resting state reported for the isolated LBD³⁸, is nevertheless more open or overextended (Supplementary Fig. 31). Binding of full agonists, such as glutamate, quisqualate or AMPA, results in closure of the clamshell by movement of domain 2 closer to domain 1 by a ~25° rotation (Supplementary Figs 30 and 31)38,77. In the context of the water-soluble LBD dimer, closure of both clamshells increases the separation of the regions linking the binding domains to the transmembrane domains by \sim 20 Å, using the α -carbon of Pro 632 (M3–S2 linker) as a reference (Supplementary Fig. 30). This movement, therefore, 'pulls apart' the M3 helices at the bundle crossing, opens the ion channel, and is the fundamental conformational change, within the binding domains, that transmits the energy associated with agonist binding to the work required to open the ion channel.

Superposition of two 'dimer' structures of the soluble LBD glutamate complex³⁸ onto the LBDs of the GluA2_{cryst} structure allows us to visualize activation-related movements in the context of the tetrameric receptor (Fig. 11 and Supplementary Fig. 32). Using α-carbon atoms of residues Lys 506, Pro 632 and Cys 773 as reference points for the S1-M1, M3-S2 and S2-M4 linkers, respectively (Supplementary Fig. 33), the most significant movement involves the M3–S2 linker, an observation in agreement with the critical role of the M3 helix in channel gating. As a consequence of the overall two-fold symmetry relating one ligand binding domain dimer to the other, the movements of the M3–S2 linkers are \sim 2-fold symmetric, that is, there are large conformational changes within dimers and smaller changes between dimers (Fig. 11c, d), consistent with chemical modification studies suggesting a breakdown in the four-fold symmetry of the ion channel upon receptor activation⁷⁸. Augmenting the intradimer motions of this simple model, there is almost certainly a perpendicular component of movement, between LBD dimers, to facilitate opening of the ~4-fold symmetric pore. A component of movement between dimers is supported by the observation that reduction of the Ile 664 disulphide bridge potentiates glutamate-induced GluA2 currents^{58,79} (Supplementary Fig. 20).

The differences in position and conformations of the LBD-TMD linkers between the A/C (proximal) and B/D (distal) subunit pairs also mean that the consequences of agonist binding and LBD clamshell closure on the TMD must necessarily be distinct (Fig. 11c, d). Evidence of this distinction can be seen by measuring the minimal distance between a reference residue within the M3-S2 linker, such as Pro 632, and the overall two-fold axis of symmetry, which is approximately coaxial with the four-fold axis of the TMD (Fig. 11c, d). In so doing, we see that the α -carbon of Pro 632 is \sim 13 Å and 25 Å distal from the two-fold axis for the A/C and the B/D subunits, respectively. Upon agonist binding and LBD closure, we predict that the α-carbons of Pro 632 will move away from the two-fold axis by ~4 Å and 7 Å for the A/C and B/D subunits, respectively. Thus, the extent of conformational movement for the proximal (A/C) and distal (B/D) subunits is substantially different and is greater for the distal (B/D) subunits. We therefore hypothesize that not only will the consequence of agonist binding to ion-channel gating be different for the proximal (A/C) and distal (B/D) subunits, but that the agonist-induced

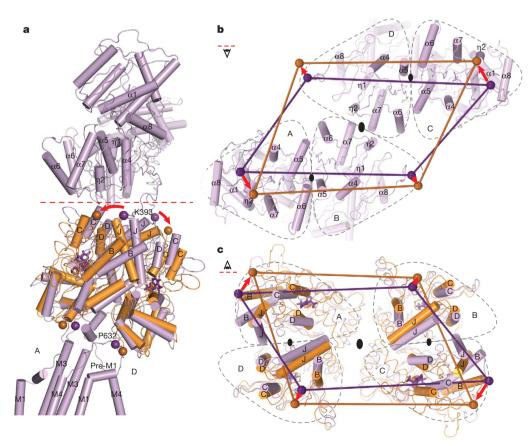


Figure 12 | iGluR desensitization. a, Structure of GluA2_{crvst} subunits A and D (purple) with superposed structure of the S729C dimer LBD (orange)58. Red dashed line indicates the interface between ATD and LBD. Purple and orange spheres indicate positions of the α-carbons for the residues Lys 393 and Pro 632. Stick models of ZK200775 and glutamate are shown in purple and orange, respectively. Red arrows indicate movement of Lys 393 during GluR desensitization. b, c, Views on the GluA2_{crvst} tetramer along the overall axis of two-fold symmetry from the interface between ATD and LBD (red dashed line in a) looking into the ATD (b) and into the LBD (c). Grey dashed lines define boundaries of the A, B, C and D subunits.

ARTICLES

conformational changes in the distal subunits may be more important to activation gating, simply because the extent of the predicted conformational change is larger. Conversely, we suggest that agonist-induced conformational changes in the proximal subunits may have a comparatively smaller role in activation gating. This explains, at least in part, why in NMDA receptors glycine binding to the GluN1 subunit, which we predict occupies the proximal position in the LBD layer and thus transmits a smaller conformational displacement to the TMD in comparison to the distal subunits, does not result in significant ion channel opening in the absence of glutamate.

Mechanism of desensitization

A hallmark of AMPA and kainate receptors is rapid and profound desensitization, or ion channel closure, after receptor activation 12,48. The molecular principles of desensitization and the structural relationships between the resting/closed and desensitized states of the ion channel pore are not yet understood for an intact receptor. To address these questions we can exploit the desensitized-like structure of the Ser 729 to Cys (S729C) mutant of the isolated LBD dimer⁵⁸. By superimposing the desensitized-like S729C dimer onto the GluA2_{cryst} structure we find that both D2 lobes of the isolated dimer superimpose well on the corresponding lobes of the tetrameric receptor, even though the isolated domains are bound with glutamate (closed clamshell) and in GluA2_{cryst}, the LBDs are 'open' and bound with antagonist. This analysis demonstrates that peptide linkers connecting the LBDs to the TMDs can adopt a similar separation in the antagonist-bound GluA2_{crvst} state and in the glutamate-bound S729C desensitized-like dimer form (Fig. 12 and Supplementary Fig. 30). Desensitization therefore simply involves rupture of the LBD D1-D1 interface and rotation of entire binding domain subunits to allow for the D2 domains and the linkers to the ion channel to adopt a closed-state-like conformation.

In the context of the GluA2_{cryst} structure, rearrangement of D1–D1 LBD interface during desensitization (Fig. 12c) demands movement of the ATDs and the ATD–LBD linkers (Fig. 12b). Upon receptor desensitization, we predict changes in both the distances between and within ATD dimers. These observations provide mechanisms, grounded in three-dimensional structure, by which binding of ions and small molecules to the ATDs of NMDA receptors can modulate receptor function⁸⁰. Ligands alter the conformation of the ATD clamshell and propagate these conformational changes throughout the receptor either directly, through the ATD–LBD linkers, or indirectly, via changes in the ATD dimer–dimer contacts across the overall two-fold axis of symmetry.

Conclusion

The GluA2_{crvst} structure uncovers domain organization and molecular symmetry of iGluRs. Crystallographic and site-directed chemical modification data demonstrate that AMPA, kainate and NMDA receptors have ~2-fold symmetric 'dimeric' extracellular domains, a ~4-fold symmetric ion channel, and a symmetry mismatch that renders diagonally related subunit pairs distinct. In heterotetrameric GluN1/ GluN2A NMDA receptors, subunits are positioned in a GluN1-GluN2A-GluN1-GluN2A pattern. Activation gating of iGluRs originates within individual LBDs with agonist binding inducing closure of the clamshell, separation of the transmembrane domain linkers, and opening of the ion channel gate. Desensitization (inactivation) results from rupture of the agonist-bound LBD dimer D1-D1 interface, leading to rigid-body rotation of individual LBD domains within the dimer and allowing for D2 domains, linkers to the ion channel and the ion channel gate to adopt a closed-state-like conformation. Gating can be perturbed at many sites by small molecule inhibitors and modulators, and by auxiliary protein subunits (Supplementary Fig. 34). The underlying architecture and symmetry of iGluRs, as revealed by the GluA2_{cryst} structure, has implications for understanding how these molecules perturb receptor function.

METHODS SUMMARY

The GluA2_{cryst}-green fluorescent protein (GFP) fusion (see Methods) was expressed in baculovirus-infected Sf9 insect cells and was purified using metalaffinity chromatography in buffers supplemented with 1.0 mM *n*-dodecyl-β-Dmaltoside. Following cleavage by thrombin to remove GFP, the receptor was further purified by size-exclusion chromatography in a buffer supplemented with 1 mM *n*-undecyl-β-D-thiomaltoside and synthetic lipids. Crystallization was performed under paraffin oil, at 4 °C, using a precipitating solution composed of 7-11% (w/v) PEG 20,000, 0.1 M 2-(N-morpholino)-ethanesulphonic acid (MES; pH 6.0–6.5) and 300 μM [[3,4-dihydro-7-(4-morpholinyl)-2,3-dioxo -6-(trifluoromethyl)-1(2H)-quinoxalinyl]methyl]phosphonic acid (ZK200775). Selenomethionine-labelled receptor was prepared and crystallized using similar conditions. Diffraction data sets were indexed, integrated and scaled using HKL2000. Initial phase information was obtained by molecular replacement using the isolated N-terminal domains and ligand-binding domains as search probes. The ion channel and interdomain polypeptides were iteratively built using the computer graphics program Coot and the structure was refined using the computer program Phenix. The ion channel functional activity of the GluA2_{cryst} construct was measured by whole-cell or outside-out patch clamp experiments. Ligand binding activity was evaluated by ³H-AMPA saturation and competition assays. NMDA receptor crosslinking experiments were done by transiently transfecting human embryonic kidney tsA 201 cells with wild-type or mutant GluN1 and GluN2A plasmid DNA and analysing oligomeric behaviour of protein on gradient SDS-PAGE gels by western blot analysis. Structure superpositions were done using Superpose (CCP4).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 August; accepted 2 November 2009. Published online 29 November 2009.

- Cowan, W. M., Sudhof, T. C. & Stevens, C. F. (eds) Synapses (The Johns Hopkins Univ. Press, 2001).
- Hayashi, T. Effects of sodium glutamate on the nervous system. Keio J. Med. 3, 183–192 (1954).
- Curtis, D. R., Phillis, J. W. & Watkins, J. C. Chemical excitation of spinal neurons. Nature 183, 611–612 (1959).
- Sugiyama, H., Ito, I. & Watanabe, M. Glutamate receptor subtypes may be classified into two major categories: A study on *Xenopus* oocytes injected with rat brain mRNA. *Neuron* 3, 129–132 (1989).
- Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. The glutamate receptor ion channels. *Pharmacol. Rev.* 51, 7–61 (1999).
- Jane, D. E., Lodge, D. & Collingridge, G. L. Kainate receptors: Pharmacology, function and therapeutic potential. *Neuropharmacology* 56, 90–113 (2009).
- Lipton, S. A. Paradigm shift in neuroprotection by NMDA receptor blockage: memantine and beyond. Nature Rev. Drug Discov. 5, 160–170 (2006).
- Alt, A., Nisenbaum, E. S., Bleakman, D. & Witkin, J. M. A role for AMPA receptors in mood disorders. Biochem. Pharmacol. 71, 1273–1288 (2006).
- Labrie, V. & Roder, J. C. The involvement of the NMDA receptor D-serine/glycine site in the pathophysiology and treatment of schizophrenia. *Neurosci. Biobehav. Rev.* doi:10.1016/j.neubiorev.2009.08.002 (18 August 2009).
- Boulter, J. et al. Molecular cloning and functional expression of glutamate receptor subunit genes. Science 249, 1033–1037 (1990).
- Keinänen, K. et al. A family of AMPA-selective glutamate receptors. Science 249, 556–560 (1990).
- 12. Sommer, B. et al. A glutamate receptor channel with high affinity for domoate and kainate. EMBO J. 11, 1651–1656 (1992).
- Moriyoshi, K. et al. Molecular cloning and characterization of the rat NMDA receptor. Nature 354, 31–37 (1991).
- Monyer, H. et al. Heteromeric NMDA receptors: molecular and functional distinction of subtypes. Science 256, 1217–1221 (1992).
- Rosenmund, C., Stern-Bach, Y. & Stevens, C. F. The tetrameric structure of a glutamate receptor channel. Science 280, 1596–1599 (1998).
- Lu, W. et al. Subunit composition of synaptic AMPA receptors revealed by a single-cell genetic approach. Neuron 62, 254–268 (2009).
- 17. Mulle, C. et al. Subunit composition of kainate receptors in hippocampal interneurons. *Neuron* **28**, 475–484 (2000).
- Christensen, J. K., Paternain, A. V., Selak, S., Ahring, P. K. & Lerma, J. A mosaic of functional kainate receptors in hippocampal interneurons. *J. Neurosci.* 24, 8986–8993 (2004).
- 19. Wo, Z. G. & Oswald, R. E. Unraveling the modular design of glutamate-gated ion channels. *Trends Neurosci.* **18**, 161–168 (1995).
- 20. O'Hara, P. J. et al. The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* 11, 41–52 (1993).
- Stern-Bach, Y. et al. Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. Neuron 13, 1345–1357 (1994).
- Wollmuth, L. P. & Sobolevsky, A. I. Structure and gating of the glutamate receptor ion channel. *Trends Neurosci.* 27, 321–328 (2004).

- Soderling, T. R. & Derkach, V. A. Postsynaptic protein phosphorylation and LTP. Trends Neurosci. 23, 75–80 (2000).
- Erreger, K., Chen, P. E., Wyllie, D. J. & Traynelis, S. F. Glutamate receptor gating. Crit. Rev. Neurobiol. 16, 187–224 (2004).
- Hansen, K. B., Yuan, H. & Traynelis, S. F. Structural aspects of AMPA receptor activation, desensitization and deactivation. Curr. Opin. Neurobiol. 17, 281–288 (2007).
- Paoletti, P. & Neyton, J. NMDA receptor subunits: function and pharmacology. Curr. Opin. Pharmacol. 7, 39–47 (2007).
- Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. Voltage-dependent block by Mg²⁺ of NMDA responses in spinal cord neurones. *Nature* 309, 261–263 (1984).
- 28. Johnson, J. W. & Ascher, P. Glycine potentiates the NMDA response in cultured mouse brain neurons. *Nature* **325**, 529–531 (1987).
- Smith, T. C. & Howe, J. R. Concentration-dependent substate behavior of native AMPA receptors. *Nature Neurosci.* 3, 992–997 (2000).
- 30. Klein, R. M. & Howe, J. R. Effects of the lurcher mutation on GluR1 desensitization and activation kinetics. *J. Neurosci.* **24**, 4941–4951 (2004).
- Vyklicky, L., Benveniste, M. & Mayer, M. L. Modulation of N-methyl-D-aspartic acid receptor desensitization by glycine in mouse cultured hippocampal neurones. J. Physiol. 428, 313–331 (1990).
- Balannik, V., Menniti, F. S., Paternain, A. V., Lerma, J. & Stern-Bach, Y. Molecular mechanisms of AMPA receptor noncompetitive antagonism. *Neuron* 48, 279–288 (2005).
- Mony, L., Kew, J. N., Gunthorpe, M. J. & Paoletti, P. Allosteric modulators of NR2Bcontaining NMDA receptors: molecular mechanisms and therapeutic potential. *Br. J. Pharmacol.* 157, 1301–1317 (2009).
- 34. Sun, Y. et al. Mechanism of glutamate receptor desensitization. *Nature* 417, 245–253 (2002).
- Jin, R. et al. Mechanism of positive allosteric modulators acting on AMPA receptors. J. Neurosci. 25, 9027–9036 (2005).
- Sobolevsky, A. I. Channel block of glutamate receptors. In Recent Research Developments in Physiology Vol. 1, Part I (ed. Pandalai, S. G.) 1–38 (Research Signpost, 2003).
- Kuusinen, A., Abele, R., Madden, D. R. & Keinänen, K. Oligomerization and ligandbinding properties of the ectodomain of the α-amino-3-hydroxy-5-methyl-4isoxazole propionic acid receptor subunit GluRD. J. Biol. Chem. 274, 28937–28943 (1999).
- Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: Crystal structures of the GluR2 ligand binding core. Neuron 28, 165–181 (2000).
- 39. Jin, R. et al. Crystal structure and association behavior of the GluR2 aminoterminal domain. EMBO J. 28, 1812–1823 (2009).
- 40. Kumar, J., Schuck, P., Jin, R. & Mayer, M. L. The N-terminal domain of GluR6-subtype glutamate receptor ion channels. *Nature Struct. Mol. Biol.* 16, 631–638 (2009).
- Tichelaar, W., Safferling, M., Keinänen, K., Stark, H. & Madden, D. R. The threedimensional structure of an ionotropic glutamate receptor reveals a dimer-ofdimers assembly. J. Mol. Biol. 344, 435–442 (2004).
- Nakagawa, T., Cheng, Y., Ramm, E., Sheng, M. & Walz, T. Structure and different conformational states of native AMPA receptor complexes. *Nature* 433, 545–549 (2005).
- Chen, G.-Q., Cui, C., Mayer, M. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* 402, 817–821 (1999).
- Mansour, M., Nagarajan, N., Nehring, R. B., Clements, J. D. & Rosenmund, C. Heteromeric AMPA receptors assemble with a preferred subunit stoichiometry and spatial arrangement. *Neuron* 32, 841–853 (2001).
- Schorge, S. & Colquhoun, D. Studies of NMDA receptor function and stoichiometry with truncated and tandem subunits. J. Neurosci. 23, 1151–1158 (2003).
- Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14, 673–681 (2006).
- 47. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11–19 (1991).
- Sommer, B. et al. Flip and flop: A cell-specific functional switch in glutamateoperated channels of the CNS. Science 249, 1580–1585 (1990).
- Turski, L. et al. ZK200775: a phosphonate quinoxalinedione AMPA antagonist for neuroprotection in stroke and trauma. Proc. Natl Acad. Sci. USA 95, 10960–10965 (1998).
- Mayer, M. L. & Armstrong, N. Structure and function of glutamate receptor ion channels. *Annu. Rev. Physiol.* 66, 161–181 (2004).
- Horning, M. & Mayer, M. Regulation of AMPA receptor gating by ligand binding core dimers. Neuron 41, 379–388 (2004).
- 52. Hollmann, M., Maron, C. & Heinemann, S. N-Glycosylation site tagging suggests a three transmembrane domain topology for the glutamate receptor GluR1. *Neuron* 13, 1331–1343 (1994).
- 53. Wo, Z. G. & Oswald, R. E. Transmembrane topology of two kainate receptor subunits revealed by N-glycosylation. *Proc. Natl Acad. Sci. USA* **91**, 7154–7158 (1994).
- Doyle, D. A. et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. Science 280, 69–77 (1998).
- Weston, M. C., Schuck, P., Ghosal, A., Rosenmund, C. & Mayer, M. L. Conformational restriction blocks glutamate receptor desensitization. *Nature Struct. Mol. Biol.* 13, 1120–1127 (2006).

- 56. Furukawa, H., Singh, S. K., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature* **438**, 185–192 (2005).
- 57. Gielen, M. et al. Structural rearrangements of NR1/NR2A NMDA receptors during allosteric inhibition. *Neuron* **57**, 80–93 (2008).
- Armstrong, N., Jasti, J., Beich-Frandsen, M. & Gouaux, E. Measurement of conformational changes accompanying desensitization in an ionotropic glutamate receptor. *Cell* 127, 85–97 (2006).
- Chang, H.-R. & Kuo, C.-C. The activation gate and gating mechanism of the NMDA receptor. J. Neurosci. 28, 1546–1556 (2008).
- Kashiwabuchi, N. et al. Impairment of motor coordination, Purkinje cell synapse formation, and cerebellar long-term depression in GluRô2 mutant mice. Cell 81, 245–252 (1995).
- 61. Zuo, J. et al. Neurodegeneration in Lurcher mice caused by mutation in $\delta 2$ glutamate receptor gene. *Nature* **388**, 769–773 (1997).
- Yelshansky, M. V., Sobolevsky, A. I., Jatzke, C. & Wollmuth, L. P. Block of AMPA receptor desensitization by a point mutation outside the ligand-binding domain. J. Neurosci. 24, 4728–4736 (2004).
- Panchenko, V. A., Glasser, C. R. & Mayer, M. L. Structural similarities between glutamate receptor channels and K⁺ channels examined by scanning mutagenesis. J. Gen. Physiol. 117, 345–360 (2001).
- 64. Kuner, T., Seeburg, P. H. & Guy, H. R. A common architecture for K⁺ channels and ionotropic glutamate receptors? *Trends Neurosci.* **26**, 27–32 (2003).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. Science 309, 897–903 (2005).
- 66. Chen, L. et al. Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature* **408**, 936–943 (2000).
- Schwenk, J. et al. Functional proteomics identify cornichon proteins as auxiliary subunits of AMPA receptors. Science 323, 1313–1319 (2009).
- Soto, D., Coombs, I. D., Kelly, L., Farrant, M. & Cull-Candy, S. G. Stargazin attenuates intracellular polyamine block of calcium-permeable AMPA receptors. *Nature Neurosci.* 10, 1260–1267 (2007).
- 69. Sobolevsky, A. I., Yelshansky, M. V. & Wollmuth, L. P. Different gating mechanisms in glutamate receptor and K+ channels. *J. Neurosci.* **23**, 7559–7568 (2003).
- Marquez-Klaka, B., Rettinger, J., Bhargava, Y., Eisele, T. & Nicke, A. Identification
 of an intersubunit cross-link between substituted cysteine residues located in the
 putative ATP binding site of the P2X₁ receptor. J. Neurosci. 27, 1456–1466 (2007).
- 71. Pedersen, S. E. & Cohen, J. B. p-Turbocurarine binding sites are located at α - γ and α - δ subunit interfaces of the nicotinic acetylcholine receptor. *Proc. Natl Acad. Sci. USA* **87**, 2785–2789 (1990).
- Kuusinen, A., Arvola, M. & Keinänen, K. Molecular dissection of the agonist binding site of an AMPA receptor. EMBO J. 14, 6327–6332 (1995).
- Armstrong, N., Sun, Y., Chen, G.-Q. & Gouaux, E. Structure of a glutamate receptor ligand binding core in complex with kainate. *Nature* 395, 913–917 (1998).
- Mayer, M. L. Crystal structures of the GluR5 and GluR6 ligand binding cores: molecular mechanisms underlying kainate receptor selectivity. *Neuron* 45, 539–552 (2005).
- 75. Inanobe, A., Furukawa, H. & Gouaux, E. Mechanism of partial agonist action at the NR1 subunit of NMDA receptors. *Neuron* 47, 71–84 (2005).
- Yao, Y., Harrison, C. B., Freddolino, P. L., Schulten, K. & Mayer, M. L. Molecular mechanism of ligand recognition by NR3 subtype glutamate receptors. *EMBO J.* 27, 2158–2170 (2008).
- Jin, R., Horning, M., Mayer, M. L. & Gouaux, E. Mechanism of activation and selectivity in a ligand-gated ion channel: structural and functional studies of GluR2 and guisqualate. *Biochemistry* 41, 15635–15643 (2002).
- Sobolevsky, A. I., Yelshansky, M. V. & Wollmuth, L. P. The outer pore of the glutamate receptor channel has 2-fold rotational symmetry. *Neuron* 41, 367–378 (2004).
- Plested, A. J. & Mayer, M. L. AMPA receptor ligand binding domain mobility revealed by functional cross linking. J. Neurosci. 29, 11912–11923 (2009).
- Gielen, M., Siegler Retchless, B., Mony, L., Johnson, J. W. & Paoletti, P. Mechanism of differential control of NMDA receptor activity by NR2 subunits. *Nature* 459, 703–707 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the personnel at beamlines 5.0.2, 8.2.1 and 8.2.2 of the Advanced Light Source and at beamline 24-ID-E of the Advanced Photon Source. We also thank T. Homrichhausen for help with cloning and FSEC screening; L. Vaskalis for assistance with illustrations; and Gouaux laboratory members for discussion. M.P.R. was supported by an individual NIH National Research Service Award. This work was supported by the NIH. E.G. is an investigator with the Howard Hughes Medical Institute.

Author Contributions All authors contributed to writing the manuscript.

Author Information Coordinates and structure factors for $GluA2_{cryst}$ and the GluA2 ligand-binding core complex bound with glutamate, LY404187 and ZK200775 have been deposited with the Protein Data Bank under codes 3KG2 and 3KGC respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

doi:10.1038/nature08624 nature

METHODS

Construct optimization and expression. The full-length rat GluA2i (flip) (NP_058957) subunit (also known as GluRBi or GluR2i)11,81, including the native signal peptide, was subcloned into the pFastBac1 vector for baculovirus expression in Sf9 insect cells using standard methods. For fluorescence detection, size-exclusion chromatography analysis (FSEC)⁴⁶ and for small- or large-scale purification, coding sequences for a thrombin cleavage site (Gly-Leu-Val-Pro-Arg-Gly), enhanced green fluorescent protein (eGFP)⁸² and an octa-histidine tag were introduced at the C terminus of each receptor construct. To improve crystallization behaviour of the GluA2 receptor, 36 residues were removed from the C terminus, 6 residues (Leu 378, Thr 379, Leu 381, Pro 382, Ser 383 and Gly 384; numbering according to the mature polypeptide sequence) were deleted from the N-terminal domain (ATD)-ligand binding domain (LBD) polypeptide linker, three of four predicted N-linked glycosylation sites were knocked out (Asn 235 to Glu; Asn 385 to Asp, and Asn 392 to Gln) and four residues in loop 1 of the LBD were substituted by alanine (Lys 410 to Ala, Glu 413 to Ala, Met 414 to Ala and Glu 416 to Ala). Point mutations of Arg 586 to Gln (Q/R site)⁴⁷ and Cys 589 to Ala were introduced to stabilize the tetrameric state of the receptor and to reduce nonspecific aggregation and disulphide bond formation, respectively. All constructs were screened by FSEC before large-scale expression experiments and crystallization trials. Together, these modifications yielded the construct, called GluA2_{crvst}, used in the crystallographic studies described here (see Supplementary Figs 1 and 2).

Purification. Infected Sf9 insect cells were harvested 72–96 h after infection and collected by a low-speed centrifugation step (6,000g, 10 min). Cells were disrupted using either an Avestin EmulsiFlex-C5 (two passes at 5,000-15,000 p.s.i.) or a Misonix Sonicator 3000 (18 × 15 s, power level 7) in a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.8 μM aprotinin, 2 μg ml⁻¹ leupeptin, 2 mM pepstatin A and 1 mM phenylmethysulphonyl fluoride (25 ml buffer/1 l Sf9 cell culture). The homogenate was clarified using a Sorval centrifuge (8,000 r.p.m., 15 min) and crude membranes were collected by ultracentrifugation (Ti45 rotor, 40,000 r.p.m., 1 h). The membranes were mechanically homogenized and subsequently solubilized for 2 h in a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 8.0 and 40 mM C₁₂M (n-dodecyl-β-D-maltopyranoside; 0.2 g of C₁₂M per 1 g membranes). Insoluble material was removed by ultracentrifugation (Ti45 rotor, 40,000 r.p.m., 40 min) and cobalt-charged metal ion affinity resin was added to the supernatant. After binding for 12-18 h, the receptor was eluted by application of buffer supplemented with 250 mM imidazole. Following thrombin digestion (1:200 mass ratio of receptor to thrombin) and concentration, the receptor containing solution was loaded onto a size-exclusion chromatography (SEC) column equilibrated in a buffer composed of 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM n-undecyl-β-D-thiomaltopyranoside (C₁₁Thio), and 0.01 mg ml⁻¹ lipid: 3:1:1 POPC:POPE:POPG (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine:1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine: 1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)]). Peak fractions were pooled and concentrated to $\sim 2 \text{ mg ml}^{-1}$ for crystallization experiments. All steps were performed at 4 °C unless otherwise noted. Analysis of the purified receptor by gel electrophoresis and tryptophan FSEC is illustrated in Supplementary Fig. 6.

Crystallization and cryoprotection. Crystallization of the rat GluA2 receptor involved multiple steps of construct modification and crystal growth optimization. In the beginning, the full-length, wild-type receptor did not yield crystals, even though the purified protein was chemically homogeneous and tetrameric in subunit stoichiometry. The first crystals, which diffracted to ~40 Å, were obtained from a construct where we introduced point mutations of Arg 586 to Gln and Cys 589 to Ala and also removed 36 residues from the C terminus. Inclusion of synthetic lipids as well as a 4-residue deletion in the ATD-LBD linker increased the diffraction limit to \sim 16 Å. The knockout of three out of four predicted N-linked glycosylation sites improved the diffraction limit to ~8 Å. Additional removal of two residues from the ATD-LBD linker, optimization of the detergent (C₁₁Thio) and ligand (ZK200775) for crystallization allowed further improvement of diffraction to 4 Å resolution. Finally, alanine substitutions in loop 1 of the LBD, continued optimization of the cryoprotection conditions and crystallization under paraffin oil allowed us to break the 4 Å diffraction limit barrier and measure crystallographic data with reasonable statistics to Bragg spacings of 3.6 Å (Supplementary Table 1).

At present, the most well ordered crystals of the GluA2 $_{cryst}$ construct grow at 4 $^{\circ}$ C in a sitting-drop configuration under paraffin oil at 4 $^{\circ}$ C. Before crystallization experiments, 0.3 mM ZK200775 or 0.3 mM ZK200775 and 0.3 mM LY404187 83 were added to the GluA2 $_{cryst}$ protein solution, which in turn was subjected to ultracentrifugation (Ti100 rotor, 40,000 r.p.m., 40 min, 4 $^{\circ}$ C). The receptor solution was then mixed with a crystallization buffer composed of 7–11% PEG 20,000 and 0.1 M MES pH 6.0–6.5 at receptor to crystallization

buffer ratio of 2:1. Crystals of the receptor appeared after 5–30 days. Cryoprotection was carried out at 4 $^{\circ}$ C by serial transfer into buffers containing increasing concentrations of glycerol, up to a maximum concentration of 30% (v/v), and then plunged into liquid nitrogen.

Measurement of complete data sets was complicated by the fact that the crystals are radiation sensitive and belong to the space group P1, with one GluA2 tetramer per asymmetric unit. The crystals have a diffraction limit of 3.5 Å and a variable mosaic spread from 0.2° to 1.5° . The unit cell dimensions are $a=91\,\text{Å},\ b=110\,\text{Å},\ c=161\,\text{Å}$ and $\alpha=85^{\circ},\ \beta=85^{\circ},\ \gamma=79^{\circ}$, yielding a Matthews coefficient $(V_{\rm M})^{84}$ of $\sim\!4.3\,\text{Å}^3\,{\rm Da}^{-1}$.

Structure determination. X-ray diffraction data sets were collected using synchrotron radiation at the Advanced Light Source (Lawrence Berkeley Laboratory; beamline 5.0.2) and were indexed, integrated and scaled using HKL200085. The structure was solved by molecular replacement using the computer program Phaser⁸⁶ and search probes composed of the rat GluA2 N-terminal domain (ATD; Protein Data Bank (PDB) code 3H5V)39 and the DNQX-bound S1S2 ligand binding domain (PDB code 1FTL)³⁸. Electron density for the ion channel and ATD-S1S2 and S1S2-channel linkers that emerged from this solution was used as a guide for building the corresponding regions of the model. The structure of the full-length GluA2_{cryst} was iteratively built and refined using the computer graphics program Coot⁸⁷ and Phenix⁸⁸, respectively (Supplementary Table 2). The four subunits in the tetrameric receptor structure contain GluA2_{crvst} residues 10 to 817, N-linked carbohydrates at Asn 349, and antagonist (ZK200775) molecules in the S1S2 ligand binding pockets. Because of weak electron density for Thr 568 to Phe 584 (helical region of M2), these residues were modelled as alanines. Regions lacking electron density and presumably disordered—residues Arg 545 to Ser 567 (M1-M2 loop), Gln 587 to Ser 592 (non-helical region of M2) and Ser 818 to Gly 832 (C terminus)—were excluded from the final GluA2_{crvst} model. Electrophysiology and ³H-AMPA binding experiments. The DNA encoding the wild-type GluA2i or GluA2_{cryst} receptors were introduced into a plasmid for expression in eukaryotic cells11 that was engineered to produce GFP via a downstream internal ribosome entry site⁴³. Human embryonic kidney tsA 201 cells grown on glass coverslips in 35-mm dishes were transiently transfected with 1–5 μg of plasmid DNA using Lipofectamine 2000 Reagent (Invitrogen). Recordings were made 24 to 48 h after transfection at room temperature. Currents from whole cells or from outside-out patches, typically held at a -60 mV potential, were recorded using an Axopatch 200B amplifier (Axon Instruments), filtered at 5 kHz and digitized at 10 kHz using HEKA Pulse software (HEKA Elektronik). The external solution contained (in mM): 140 NaCl, 2.4 KCl, 4 CaCl₂, 4 MgCl₂, 10 HEPES pH 7.3 and 10 glucose; 10 mM NaCl and 10 mM sucrose were added to the extracellular activating solution containing 1 mM L-glutamate. The internal solution contained (in mM): 150 CsCl, 10 NaCl, 10 EGTA, 20 HEPES pH 7.3. Rapid solution exchange was achieved with a twobarrel theta glass pipette controlled by a piezoelectric translator; junction currents were used to estimate speed of solution exchange after recordings. Data analysis was performed using Origin.

Binding experiments with ³H-AMPA were performed following the previously published protocol⁸⁹.

Cysteine crosslinking and western blots. Single cysteine substitutions were introduced using conventional PCR-based methods. Constructs were verified by sequencing over the entire length of the iGluR coding region. The parent GluA2 $_{\rm cryst}$ or single cysteine substituted constructs in the pFastBac1 vector were expressed in Sf9 insect cells and purified as described above. About 5 μg of protein was subjected to denaturing conditions by addition of 6× SDS sample buffer containing 300 mM Tris-H:Cl (pH 6.8), 12% SDS, 0.6% bromophenol blue and 60% glycerol in the absence (non-reducing condition) or presence (reducing condition) of 100 mM dithiothreitol (DTT). The protein samples were then run on a 4–15% gradient SDS–PAGE gel and protein bands were visualized by Coomassie blue staining.

For NMDA receptor crosslinking experiments, DNA encoding wild-type rat GluN1a (EDL93606) and GluN2A (NP_036705) subunits were introduced into plasmids for expression in eukaryotic cells. The same amounts of wild-type or mutant GluN1 and GluN2A plasmid DNA (1–5 µg) mixed together and Lipofectamine 2000 Reagent (Invitrogen) were used to transiently transfect human embryonic kidney tsA 201 cells. Cells were harvested 24–48 h after transfection and solubilized for 1 h in PBS supplemented with 40 mM $C_{12}M$ ($n\text{-}dodecyl-\beta\text{-}D\text{-}maltopyranoside}$), 0.8 µM aprotinin, 2 µg ml $^{-1}$ leupeptin, 2 mM pepstatin A and 1 mM phenylmethysulphonyl fluoride. Crude solubilized material was clarified by ultracentrifugation (Ti100 rotor, 40,000 r.p.m., 40 min, 4 $^{\circ}$ C) and then run on a 4–15% gradient SDS–PAGE gel either in the absence (non-reducing condition) or presence (reducing condition) of 100 mM DTT. Protein bands were transferred to the Hybond-ECL nitrocellulose membranes (Amersham Biosciences) using Trans-Blot electrophoretic transfer cell (Bio-Rad). The membranes were blocked for 1 h at room temperature in TBST (150 mM NaCl,

doi:10.1038/nature08624 nature

10 mM Tris-HCl pH7.6, 0.1% Tween-20) containing 5% non-fat milk and 1% BSA and then incubated for 1 h with either mouse anti-NMDAR1 monoclonal (Millipore, MAB1586) or rabbit anti-NMDAR2A polyclonal (Millipore, AB1555P) antibodies. After four 15-min washes with TBST, the membranes were incubated for 1 h at room temperature with either sheep anti-mouse (GE Healthcare, NA931V) or donkey anti-rabbit (GE Healthcare, NA934V) antibodies conjugated to horseradish peroxidase. Then the membranes were washed again four times for 15 min with TBST and immunoreactivity was visualized using the ECL detection kit (Amersham Biosciences).

Structure superpositions. The structures were superimposed using the CCP4 program Superpose. The ATD dimer-of-dimers from GluA2_{cryst} (ABCD subunits) and ATD dimer-of-dimers represented by crystal packing of the isolated GluK2 ATD⁴⁰ (ABA'B' subunits) were superposed based on α -carbon atoms for 292 residues per subunit in the structurally conserved regions (overall sequence identity of the regions, 25.5%): residues 4–14, 17–31, 34–93, 102–109, 111–131, 140–159, 171–259, 273–293, 320–341 and 351–375 for GluA2_{cryst} and 3–13, 20–34, 42–101, 113–120, 125–145, 154–173, 180–268, 286–306, 325–346 and 357–381 for GluK2, respectively. The ion channels of GluA2_{cryst} (ABCD subunits) and KcsA (ADCB subunits) were superposed based on α -carbon atoms for 64 residues per subunit in the structurally conserved regions (overall sequence identity of the regions, 20%): residues 522–544, 575–584 and 595–625 for GluA2_{cryst} and 29–51, 61–70 and 85–115 for KcsA, respectively. The ion channels of Shaker and KcsA were

superposed based on α -carbon atoms of the P-loop region (residues 358–375 in Shaker and 63–80 in KcsA, respectively) yielding the r.m.s.d. of 0.7 Å.

- Hollmann, M., O'Shea-Greenfield, A., Rogers, S. W. & Heinemann, S. Cloning by functional expression of a member of the glutamate receptor family. *Nature* 342, 643–648 (1989).
- 82. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173, 33–38 (1996).
- Miu, P. et al. Novel AMPA receptor potentiators LY392098 and LY404187: effects on recombinant human AMPA receptors in vitro. Neuropharm. 40, 976–983 (2001).
- Matthews, B. W. Solvent content of protein crystals. J. Mol. Biol. 33, 491–497 (1968).
- 85. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- 86. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D.* **63**, 32–41 (2007).
- 87. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D.* **60**, 2126–2132 (2004).
- 88. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D. 58, 1948–1954 (2002).
- 89. Chen, G. Q. & Gouaux, E. Overexpression of a glutamate receptor (GluR2) ligand binding domain in *Escherichia coli*: Application of a novel protein folding screen. *Proc. Natl Acad. Sci. USA* **94**, 13431–13436 (1997).

ARTICLES

Crystal structure of a bacterial homologue of the kidney urea transporter

Elena J. Levin¹, Matthias Quick^{2,3} & Ming Zhou¹

Urea is highly concentrated in the mammalian kidney to produce the osmotic gradient necessary for water re-absorption. Free diffusion of urea across cell membranes is slow owing to its high polarity, and specialized urea transporters have evolved to achieve rapid and selective urea permeation. Here we present the 2.3 Å structure of a functional urea transporter from the bacterium *Desulfovibrio vulgaris*. The transporter is a homotrimer, and each subunit contains a continuous membrane-spanning pore formed by the two homologous halves of the protein. The pore contains a constricted selectivity filter that can accommodate several dehydrated urea molecules in single file. Backbone and side-chain oxygen atoms provide continuous coordination of urea as it progresses through the filter, and well-placed α -helix dipoles provide further compensation for dehydration energy. These results establish that the urea transporter operates by a channel-like mechanism and reveal the physical and chemical basis of urea selectivity.

Urea is ubiquitous in nature. Bacteria take up urea and convert it to ammonia for use as a nitrogen source, and in certain enteric pathogens, a buffer for surviving the extreme acidic conditions in the stomach^{1,2}. In higher organisms such as mammals, urea is produced as an end product of protein catabolism because it is less toxic than ammonia and more soluble than uric acid. As well as being a vehicle for nitrogen excretion, urea is used as an osmolyte. For example, sharks and rays use urea to maintain isosmotic with sea water³, and mammals are able to concentrate urea by hundreds of fold in the kidney to provide the osmotic gradient essential for water re-absorption^{4,5}.

Urea has a stronger dipole moment (4.6 D) than water (1.8 D), and its unassisted diffusion across lipid bilayers is slow⁶. There are at least four families of transporters that facilitate selective permeation of urea: an ATP-dependent ABC type urea transporter⁷; an ion-motive force-dependent urea transporter⁸; an acid-activated urea channel that belongs to the urea/amide channel family²; and finally the urea transporter (UT) family, which is the most widely distributed family and the focus of this research. Since the first cloning of a UT from mammalian kidney by expression cloning⁹, UT members have been found in bacteria, fungi, insects and many vertebrates including all mammals.

Previous characterizations of mammalian and bacterial UTs^{1,9-12} have demonstrated that they facilitate the diffusion of urea and urea analogues along their concentration gradients at rates between 10⁴ and 10⁶ s⁻¹ (refs 10, 13), consistent with a channel-like mechanism. In the absence of structural information, how urea transporters coordinate and stabilize an at least partially dehydrated urea to achieve selectivity and facilitate permeation remains unknown. To address these questions, we focused on the functional and structural characterization of a UT from the bacterium D. vulgaris, dvUT, which is homologous to mammalian UTs (Supplementary Fig. 1a). All UT sequences have two homologous halves that probably arose from the duplication of an ancient gene¹⁴, and are predicted to contain ten transmembrane helices. Identity between the transmembrane domains of dvUT and human UTs is around 35%, and clusters of highly conserved residues are distributed throughout the primary sequence. We demonstrate that dvUT is a functional urea

transporter, and present the atomic resolution structure of dvUT and the molecular basis it reveals for selective permeation of urea.

Functional characterization of dvUT

The function of dvUT was examined in two assays. First, urea flux through dvUT was measured in a tracer uptake assay^{2,9}. dvUT was expressed in *Xenopus laevis* oocytes by injection of complimentary RNA, and the uptake of ^{14}C -labelled urea ($\sim 181~\mu\text{M})$ by individual oocytes was measured. The uptake of labelled urea increased with time and reached $\sim 50~\text{pmol}$ per oocyte in about 60 min (Fig. 1a). Assuming the volume of an oocyte is $\sim 0.25~\mu\text{l}$ (ref. 9), the estimated intracellular concentration of labelled urea is $\sim 200~\mu\text{M}$, essentially in equilibrium with the extracellular urea. In contrast, urea accumulation in water-injected oocytes showed a much slower time course and had not reached equilibrium even after 120 min (Fig. 1a).

Because urea flux through mammalian UTs is inhibited by phloretin^{9,15}, we examined its effect on both human UT-B and dvUT (Fig. 1b). The uptake of urea by oocytes expressing UT-B reached equilibrium in ~15 min, comparable to a previous report⁹ and faster than oocytes expressing dvUT. However, this does not necessary indicate that dvUT has a slower flux rate because the relative amounts of the two functional UT proteins expressed on the membrane are unknown. Phloretin (1 mM) inhibited facilitated urea transport through UT-B and dvUT (Fig. 1b) by 86% and 87%, respectively, suggesting a similar mode of interaction of phloretin with both proteins. In fact, phloretin also inhibits a urea transporter from *Actinobacillus pleuropneumoniae* (apUT)¹¹, indicating that a common architecture is probably shared by prokaryotic and eukaryotic UTs.

Second, equilibrium binding of urea to dvUT was measured in a scintillation proximity assay (SPA)^{16,17}. In this assay, purified Histagged dvUT was immobilized on copper-coated scintillation beads that emit light only when a radioactive ligand stably binds to dvUT. The addition of ¹⁴C-labelled urea induced light emission from the scintillant, indicating urea binding to dvUT. Fitting the data to the Hill equation revealed an apparent equilibrium dissociation constant (K_d) of 2.3 \pm 0.14 mM (mean \pm standard error) (Fig. 1c) with a Hill coefficient of 3.4 \pm 0.7, indicative of cooperative binding of about

¹Department of Physiology & Cellular Biophysics, ²Department of Psychiatry, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032, USA. ³Division of Molecular Therapeutics, New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York 10032, USA.

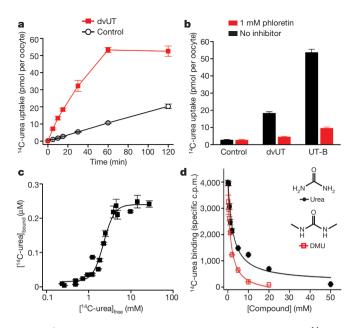


Figure 1 | **dvUT** mediated urea flux and binding. **a**, Time course of ¹⁴C-urea uptake in oocytes injected with dvUT cRNA (squares) or water (circles). **b**, Uptake of radiolabelled urea in the presence (red) or absence (black) of 1 mM phloretin. **c**, Saturation equilibrium urea binding by dvUT. The solid line represents fitting of the data to the Hill equation. **d**, SPA-based ¹⁴C-urea equilibrium binding in the presence of increasing concentrations of urea (circles) and DMU (squares). The solid lines correspond to data fit with a single-site binding isotherm. c.p.m., counts per minute. Error bars in all panels are s.e.m. of 3–10 measurements.

three urea molecules per molecule of dvUT, and suggesting that dvUT has several urea-binding sites.

Interaction of N,N'-dimethylurea (DMU)—a urea analogue that was used in the structural studies—with dvUT was measured in an SPA-based competition assay (Fig. 1d). The addition of increasing concentrations of DMU to dvUT incubated with a fixed concentration of labelled urea caused progressive loss of light emission, indicating that the bound urea was displaced. The concentration of DMU necessary to achieve half-maximum inhibition (IC50) was 1.4 ± 0.2 mM. As a control experiment, unlabelled urea was used for isotopic dilution of the 14 C-labelled urea. A 50% reduction of 14 C-urea binding was obtained at 2.3 ± 0.7 mM urea, consistent with the $K_{\rm d}$ measured in the previous equilibrium binding assay (Fig. 1c) and slightly higher than the IC50 of DMU (Fig. 1d). These results indicate that DMU and urea may occupy similar sites in dvUT.

Channel architecture

The structure of dvUT was determined by single-wavelength anomalous dispersion (SAD) using a mercury-derivatized crystal, and refined to 2.3 Å using a gold-derivatized crystal (Methods and Supplementary Table 1). The final structure contains one protomer in the asymmetric unit, which has residues 1–163 and 168–334 of dvUT along with 9 gold atoms and 55 waters. Three residues from the carboxy terminus and four residues from a long loop connecting the homologous halves of the protein are disordered and omitted from the model.

The dvUT protomer contains two hemi-cylindrical domains of six helices each, and the two domains are related by a rotational pseudotwo-fold symmetry axis lying in the plane of the membrane (Fig. 2a and Supplementary Fig. 1b). The first helix of each domain, which we call pore helix a or b (Pa or Pb), is tilted at a roughly 50° angle with respect to the membrane norm and extends $\sim\!10\,\text{Å}$ into the membrane. The pore helices end in loops that turn sharply to exit on the same side of the membrane. The next four helices from each domain, T1a–T4a and T1b–T4b, span the entire membrane. The remaining helices, T5a and T5b, are perpendicular to the membrane and unwind at the middle of the membrane into an extended coil to cross the membrane. The α -carbon atoms of the first half, Pa and T1a–T5a, can be superposed onto the second half, Pb and T1b–T5b, by a rotation of 179.5° with a root mean square deviation (r.m.s.d.) of 1.13 Å.

The amino and C termini exit the membrane on the same side of the protein and both end in short helical segments (Fig. 2a and Supplementary Fig. 1b). Although the orientation of dvUT in the plasma membrane has not been experimentally determined, the excess of positive charge on the termini-containing face suggests that this side orients towards the cytoplasm¹⁸, and this orientation is consistent with the known topology of its mammalian homologues^{19–21}.

dvUT crystallizes as a homotrimer (Fig. 2b), with its three-fold rotational axis coincident with a crystallographic symmetry axis. Helices T4a and T5a of one subunit interact with helices T4b and T5b of the neighbouring subunit. The interface between the subunits is largely hydrophobic with a total interacting surface area of 5,010 Ų. On the three-fold symmetry axis is a large cavity (Supplementary Fig. 2a). Both ends of the cavity are sealed off from the bulk solvent by two layers of tightly packed hydrophobic side chains circling the three-fold rotational axis: Leu 160 and Pro 287 on the periplasmic face, and Phe 120 and Trp 124 on the cytoplasmic face (Supplementary Fig. 2b, c). The residues lining the walls of the central cavity are largely hydrophobic; tryptophan, phenylalanine and leucine residues predominate. Several large electron density

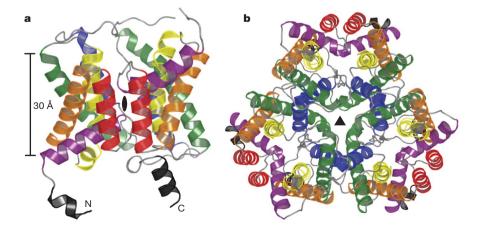


Figure 2 | **Fold and oligomeric structure of dvUT. a**, Cartoon representation of the dvUT protomer. The two-fold pseudo-symmetry axis, marked as a black oval, is normal to the plane of the figure. Colour of helices matches that

in the topology diagram (Supplementary Fig. 1b). **b**, Cartoon representation of the full dvUT trimer. The crystallographic three-fold symmetry axis is marked as a black triangle.

peaks in the cavity probably correspond to partially ordered detergent or lipid molecules.

Several lines of evidence suggest that the observed trimer is not an artefact of crystal packing. In addition to the large area of the interface, chemical crosslinking experiments support that detergent-solubilized dvUT is a trimer (Supplementary Fig. 3a); also, the same homotrimer was observed in a lower resolution structure obtained from the native protein, which crystallizes in a lower symmetry space group with different packing (Supplementary Fig. 3b, c).

Selectivity filter

Each dvUT protomer has a continuous solvent accessible permeation pathway that is formed between the two homologous halves of the protein (Fig. 3a). This pore has a constricted region, ~16 Å long, that opens into two wide vestibules on either side. We define the constricted region as the selectivity filter. Highly conserved residues from six different regions of the protein, Pa and Pb, T3a and T3b, and T5a and T5b, are brought together to form the selectivity filter (Supplementary Fig. 6a). The residues that line the selectivity filter are coloured in Supplementary Fig. 1a and shown in Fig. 3b. One side of the selectivity filter has two linear arrays of three oxygen atoms, which we call the oxygen ladders (Fig. 3b). Each oxygen ladder is flanked by two parallel and closely spaced phenylalanine side chains that compress the filter into a slot-like shape. Hydrophobic phenylalanine and leucine residues complete the lining of the filter opposite to each of the oxygen ladders. Between the two oxygen ladders, there is a gap $(\sim 6 \text{ Å})$ packed by two valine side chains, Val 25 and Val 188. A pair of leucine side chains, Leu 84 and Leu 247, constricts this part of the filter just like the phenylalanines above and below, and on the opposite side of the valines are two threonine side chains, Thr 130 and Thr 294. Despite the relatively weak overall conservation between the two halves of the protein, the nature and positioning of the side chains forming the wall of the selectivity filter are remarkably symmetrical: with the exception of a single Gln 24/Glu 187 pair, every residue in the filter has an identical symmetry-related partner (Fig. 3b). For each oxygen ladder, the distances between the oxygen

atoms are 3.4–3.6 Å, except for the distance between the side chain and backbone oxygens of Gln 24, which is 4 Å. However, the position of this side chain may be perturbed in the crystal structure by its interaction with a gold atom, added during the crystallization process, sitting at the entrance of the pore (Fig. 3b).

The permeability of dvUT to water has not been determined experimentally; however, water permeation has been observed in apUT¹¹ and a mammalian UT²², although there is controversy over the latter¹⁰. In the absence of ligand, three positive electron density peaks are visible in selectivity filter, which probably correspond to partially ordered water molecules and raise the possibility that dvUT is water permeable. Two of the peaks lie directly between the pairs of phenyalanines and are within hydrogen-bonding distance of the oxygen ladders. The third is hydrogen-bonded to the side-chain hydroxyl of Thr 294.

Potential urea-binding sites

To investigate how selective permeation of urea is facilitated by the physical and chemical properties of the selectivity filter, we cocrystallized dvUT with urea. After initial attempts to obtain high-resolution data in the presence of urea were unsuccessful, we co-crystallized dvUT with DMU. A 2.4 Å data set was collected on a crystal grown in the presence of 10 mM DMU.

The structure of dvUT–DMU was solved by molecular replacement, and two strong electron densities were observed in a difference map. Each density is flat and triangular in shape, and is located close to an oxygen ladder (Fig. 3c, e and stereo image in Supplementary Fig. 4). The triangle-shaped flat electron density makes the orientation of DMU unambiguous, with its two amide nitrogen atoms facing an oxygen ladder. We define the two sites as S_o and S_i, for their proximity to the outside or inside of a cell (Fig. 3b).

We propose that a partially dehydrated DMU in the S_o site is stabilized by four different interactions. First, the two amide nitrogen atoms on DMU are 2.9 and 2.6 Å away from the side-chain and backbone oxygen atoms of Glu 187, respectively, poised to form hydrogen bonds. Second, the two side chains of Phe 190 and Phe 243 sandwich DMU so that the two aromatic rings provide stabilization of the partial

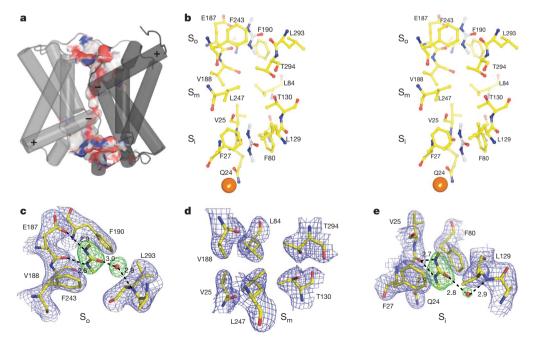


Figure 3 | **Structure of the dvUT pore and DMU-binding sites. a**, A surface representation of the dvUT pore with T1a, T1b and the N and C termini removed. Oxygen and nitrogen atoms are coloured in red and blue, respectively. **b**, Stereo view of residues lining the selectivity filter. Two DMU molecules, coordinates taken from the dvUT–DMU complex, are shown in the S_o and S_i sites. A gold atom that co-crystallizes with dvUT is shown in

gold. **c–e**. Views of the S_o (**c**), S_m (**d**) and S_i (**e**) regions of the selectivity filter for dvUT–DMU complex. The dark blue mesh corresponds to $2F_o-F_c$ electron density map contoured at 1.5σ . The green mesh in the S_o and S_i sites corresponds to 3.0σ F_o-F_c electron density calculated with DMU and the displayed water molecule omitted.

positive charge on the amide nitrogen by amide— π stacking interactions²³. Third, the carbonyl oxygen atom on DMU is 3.0 Å away from a water molecule that in turn hydrogen bonds with a backbone nitrogen atom (Fig. 3c). Fourth, the oxygen ladder is at the very end of Pb (Figs 3a and 4), and this arrangement conveniently uses the partial negative charge created by the α -helix dipole to further stabilize the partial positive charge on the amide nitrogens. The use of helix dipoles to stabilize bound permeants has been noted previously in potassium channels²⁴ and in chloride transporters²⁵.

Binding of DMU in the S_i site is similar to that at the S_o site, but shifted a few tenths of an angstrom towards the centre of the membrane so that both of the amide nitrogen atoms are 2.7 and 2.8 Å away from the centre oxygen in the ladder (Fig. 3e). The slight shift is probably because the side-chain oxygen atom of Gln 24 is not available because it is farther away to coordinate a nearby gold atom.

Given the similarities in shape and electrostatics between urea and DMU, urea probably occupies the S_o and S_i sites in a similar manner as DMU. We therefore modelled urea into the two sites using the DMU coordinates (Fig. 4). For urea to complete its journey across the cell membrane, it must pass through the space between S₀ and S_i, which we define as the S_m site because it is in the middle of the membrane (Fig. 3b, d). We modelled a urea molecule into this site by positioning the two amide nitrogen atoms within hydrogenbonding distance to the two innermost oxygen atoms from the oxygen ladders (Fig. 4). This arrangement naturally places the carbonyl oxygen atom of urea within hydrogen-bonding distance to the hydroxyls from both Thr 130 and Thr 294-two residues that are 100% conserved in all known UT sequences. Unlike urea in the S_o and S_i sites, a urea in S_m is sandwiched by two leucine side chains ~6.1 Å apart that also exhibit strong conservation in mammalian UTs. We propose that the S_m site in the middle of the selectivity filter is superbly suited to screen a permeating ligand for the appropriate electrostatic properties as well as a stringent test of planarity. This site however, is probably not accessible to DMU because the methyl groups will clash with the two threonine side chains.

Summary

The structure of dvUT showed a continuous solvent accessible pore in the structure, indicating that dvUT, and by extrapolation, UTs in general probably operate by a channel-like mechanism. Selective permeation of urea is achieved by a long and narrow selectivity filter that allows dehydrated urea molecules to permeate in single file. Compensation for dehydration energy is provided by continuous coordination of urea with hydrogen bonds as it goes through the selectivity filter, and by well-placed α -helix dipoles and amide- π

interactions (Fig. 4). Furthermore, as a urea molecule progresses through the filter, its orientation is rigorously maintained by closely spaced hydrophobic residues, so that optimal hydrogen bonding and single-file conduction can be achieved. This mechanism is consistent with the ability of UTs to exclude molecules larger than urea, but allow smaller analogues such as formamide to pass through.

It is not immediately apparent how universal the trimeric complex is among urea transporters. A recent study on apUT argued that that particular homologue was a dimer¹¹. The stoichiometry of mammalian UT remains unknown. The existence of the mammalian UT-A1 isoform, which consists of two tandem copies of UT sequences, suggests that UT-A1 forms oligomers with even numbers of UT domains. However, UT-A1 co-expresses in the same cells with UT-A3 (ref. 26), which only has one UT domain, thus raising the possibility that UT-A1 and UT-A3 assemble to form a trimer-like complex, although attempts to isolate a UT-A1-UT-A3 complex by co-immunoprecipitation have not been successful²⁷. Given that UTs seem to conduct urea constitutively, and that each subunit contains its own complete pore, it is entirely possible that differences in native oligomeric state between homologues would not have a large effect on their basic function.

Similar to other channels that conduct neutral molecules, such as ammonia^{28,29}, water^{30,31} or glycerol³², UT is constructed by two oppositely oriented homologous halves. The arrangement of the UT transmembrane helices T1a-T5a and T1b-T5b bears a resemblance to the fold of the AMT (ammonium transporter)/Rh (rhesus) family of ammonia channels^{28,29} (Supplementary Fig. 5a, b). Together with the weak sequence identity between the two proteins (22% between dvUT and Escherichia coli AmtB), this would suggest that the two families are evolutionarily related. However, there is little similarity between the two proteins in the structure of the permeation pathway itself. The most significant difference is that the tilted helices Pa and Pb that contribute the residues forming the oxygen ladders are not present in AmtB. Their place is taken in the AmtB structure by mostly hydrophobic side chains from helices T1 and T6 (equivalent to T1a and T1b in dvUT, which are not involved in the pore) (Supplementary Fig. 6a, b). The selectivity filter of AmtB is narrower overall than dvUT (Supplementary Fig. 6c), and is largely hydrophobic except for a pair of central histidine residues that invite comparison to Thr 130 and Thr 294 in dvUT (Supplementary Fig. 6d, e). Furthermore, dvUT lacks the 11th transmembrane segment possessed by AmtB, and the trimer interface is constructed by different transmembrane helices (Supplementary Fig. 5c, d).

The structure of dvUT has revealed the physical and chemical principles governing selective permeation of urea in UTs. The structure also provides a framework for further studies to understand how

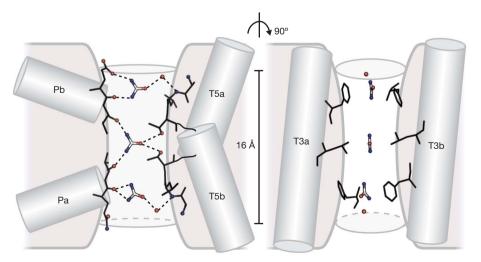


Figure 4 | **Schematic view of the selectivity filter.** The selectivity filter is shown from two angles. The predicted locations of three urea molecules and their hydrogen-bonding partners are on the left. In the perpendicular

direction, the filter is compressed by phenylalanine and leucine side chains lining the walls of the pore (right). Helices contributing residues to the selectivity filter are represented as grey cylinders.

the rate of urea permeation is determined, and how the regulation of UT function, non-covalently by small-molecule compounds^{9,33} and covalently by phosphorylation^{34,35}, is achieved.

METHODS SUMMARY

dvUT was identified as a suitable homologue for structural studies in an initial expression screen conducted by the New York Consortium on Membrane Protein Structure (NYCOMPS, see Methods), and then adapted to the Smt3 system (Invitrogen) for enhanced yield. An N-terminal His-tagged SUMOdvUT fusion protein was expressed in E. coli BL21(DE3) cells and purified on a cobalt affinity column. After cleavage of the affinity tag and SUMO domain with SUMO protease, and a second round of chromatography on a Superdex 200 10/300 GL gel filtration column, the protein was concentrated to ~ 8 mg ml⁻¹ in buffer containing 300 mM NaCl, 20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol and 40 mM *n*-octyl-β-D-maltoside (OM). Crystals were obtained by sittingdrop vapour diffusion in mother liquor containing 22% PEG1500, 100 mM sodium cacodylate, pH 6.5 and 10 mM thiomersal or potassium gold cyanide. A 2.5 Å mercury-derivatized and a 2.3 Å gold-derivatized data set were collected on beamlines X25 and X29 at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. The mercury data set was used for location of heavy atom sites and SAD phasing. After density modification and automated and manual model building, the resulting partial structure was used as a molecular replacement model for the gold data set. Iterative rounds of building and refinement produced a final model with 330 out of 337 residues resolved and R and R_{free} values of 17.9 and 20.4%, respectively. High resolution DMU-dvUT crystals were obtained by co-crystallizing the protein with 10 mM DMU as well as potassium gold cyanide, and the 2.4 Å ligand-bound structure was solved by molecular replacement. For the SPA binding assays, $100\,\mu$ l of 2.5 mg ml⁻¹ Cu²⁺-coated YSi SPA beads (GE Healthcare), $181\,\mu$ M 14 C-urea (55 mCi mmol⁻¹), 50–250 ng dvUT, and appropriate concentrations of cold urea or DMU were combined per assay in clear bottomed/white-wall 96-well plates in buffer containing 150 mM Tris/Mes, pH 7.5 with 50 mM NaCl, 20% glycerol, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), and 0.1% n-dodecyl-β-D-maltoside (DDM). Reactions performed in the presence of 400 mM imidazole served as negative controls for all conditions tested.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 June; accepted 8 October 2009. Published online 28 October 2009.

- Sebbane, F. et al. The Yersinia pseudotuberculosis Yut protein, a new type of urea transporter homologous to eukaryotic channels and functionally interchangeable in vitro with the Helicobacter pylori Urel protein. Mol. Microbiol. 45, 1165–1174 (2002).
- Weeks, D. L., Eskandari, S., Scott, D. R. & Sachs, G. A H⁺-gated urea channel: the link between *Helicobacter pylori* urease and gastric colonization. *Science* 287, 482–485 (2000).
- 3. Hediger, M. A. et al. Structure, regulation and physiological roles of urea transporters. *Kidney Int.* **49**, 1615–1623 (1996).
- Sands, J. M. Mammalian urea transporters. Annu. Rev. Physiol. 65, 543–566 (2003).
- Bagnasco, S. M. Role and regulation of urea transporters. Pflugers Arch. 450, 217–226 (2005).
- Finkelstein, A. Water and nonelectrolyte permeability of lipid bilayer membranes. J. Gen. Physiol. 68, 127–135 (1976).
- Valladares, A., Montesinos, M. L., Herrero, A. & Flores, E. An ABC-type, highaffinity urea permease identified in cyanobacteria. *Mol. Microbiol.* 43, 703–715 (2002).
- 8. Kojima, S., Bohner, A. & von Wiren, N. Molecular mechanisms of urea transport in plants. *J. Membr. Biol.* **212**, 83–91 (2006).
- You, G. et al. Cloning and characterization of the vasopressin-regulated urea transporter. Nature 365, 844–847 (1993).
- MacIver, B., Smith, C. P., Hill, W. G. & Zeidel, M. L. Functional characterization of mouse urea transporters UT-A2 and UT-A3 expressed in purified *Xenopus laevis* oocyte plasma membranes. *Am. J. Physiol. Renal Physiol.* 294, F956–F964 (2008).
- Raunser, S. et al. Oligomeric structure and functional characterization of the urea transporter from Actinobacillus pleuropneumoniae. J. Mol. Biol. 387, 619–627 (2009).
- Zhao, D., Sonawane, N. D., Levin, M. H. & Yang, B. Comparative transport efficiencies of urea analogues through urea transporter UT-B. *Biochim. Biophys.* Acta 1768, 1815–1821 (2007).
- Mannuzzu, L. M., Moronne, M. M. & Macey, R. I. Estimate of the number of urea transport sites in erythrocyte ghosts using a hydrophobic mercurial. *J. Membr. Biol.* 133, 85–97 (1993).

- Minocha, R., Studley, K. & Saier, M. H. Jr. The urea transporter (UT) family: bioinformatic analyses leading to structural, functional, and evolutionary predictions. Receptors Channels 9, 345–352 (2003).
- Chou, C. L. & Knepper, M. A. Inhibition of urea transport in inner medullary collecting duct by phloretin and urea analogues. Am. J. Physiol. 257, F359–F365 (1989).
- Quick, M. & Javitch, J. A. Monitoring the function of membrane transport proteins in detergent-solubilized form. Proc. Natl Acad. Sci. USA 104, 3603–3608 (2007).
- Shi, L., Quick, M., Zhao, Y., Weinstein, H. & Javitch, J. A. The mechanism of a neurotransmitter:sodium symporter—inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol. Cell* 30, 667–677 (2008).
- 18. von Heijne, G. & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* 174, 671–678 (1988).
- 19. Shayakul, C., Steel, A. & Hediger, M. A. Molecular cloning and characterization of the vasopressin-regulated urea transporter of rat kidney collecting ducts. *J. Clin. Invest.* **98**, 2580–2587 (1996).
- Bradford, A. D. et al. 97- and 117-kDa forms of collecting duct urea transporter UT-A1 are due to different states of glycosylation. Am. J. Physiol. Renal Physiol. 281, F133-F143 (2001).
- 21. Lucien, N. et al. Antigenic and functional properties of the human red blood cell urea transporter hUT-B1. J. Biol. Chem. 277, 34101–34108 (2002).
- Yang, B. & Verkman, A. S. Analysis of double knockout mice lacking aquaporin-1 and urea transporter UT-B. Evidence for UT-B-facilitated water transport in erythrocytes. J. Biol. Chem. 277, 36782–36786 (2002).
- Imai, Y. N., Inoue, Y., Nakanishi, I. & Kitaura, K. Amide–π interactions between formamide and benzene. J. Comput. Chem. 30, 2267–2276 (2009).
- Doyle, D. A. et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. Science 280, 69–77 (1998).
- Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a CIC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* 415, 287–294 (2002).
- Terris, J. M., Knepper, M. A. & Wade, J. B. UT-A3: localization and characterization of an additional urea transporter isoform in the IMCD. Am. J. Physiol. Renal Physiol. 280, F325–F332 (2001).
- Blount, M. A., Klein, J. D., Martin, C. F., Tchapyjnikov, D. & Sands, J. M. Forskolin stimulates phosphorylation and membrane accumulation of UT-A3. Am. J. Physiol. Renal Physiol. 293, F1308–F1313 (2007).
- Khademi, S. et al. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. Science 305, 1587–1594 (2004).
- Zheng, L., Kostrewa, D., Berneche, S., Winkler, F. K. & Li, X. D. The mechanism of ammonia transport based on the crystal structure of AmtB of Escherichia coli. Proc. Natl Acad. Sci. USA 101, 17090–17095 (2004).
- Murata, K. et al. Structural determinants of water permeation through aquaporin-1. Nature 407, 599–605 (2000).
- Sui, H., Han, B. G., Lee, J. K., Walian, P. & Jap, B. K. Structural basis of waterspecific transport through the AQP1 water channel. *Nature* 414, 872–878 (2001).
- 32. Fu, D. et al. Structure of a glycerol-conducting channel and the basis for its selectivity. Science 290, 481–486 (2000).
- 33. Levin, M. H., de la Fuente, R. & Verkman, A. S. Urearetics: a small molecule screen yields nanomolar potency inhibitors of urea transporter UT-B. *FASEB J.* 21, 551–563 (2007).
- 34. Shayakul, C. & Hediger, M. A. The SLC14 gene family of urea transporters. *Pflugers Arch.* 447, 603–609 (2004).
- Zhang, C., Sands, J. M. & Klein, J. D. Vasopressin rapidly increases phosphorylation of UT-A1 urea transporter in rat IMCDs through PKA. Am. J. Physiol. Renal Physiol. 282, F85–F90 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. MacKinnon for advice and support throughout the project, C. Miller and E. Gouaux for comments on the manuscript, J. Love and NYCOMPS for cloning and the initial screen of protein expression levels, Y. Pan for messenger RNA preparation and oocyte injection, and J. Weng and Y. Cao for crystal screening and data collection at the synchrotrons. Data for this study were measured at beamlines X4A, X4C, X25 and X29 of the NSLS and the NE-CAT 24ID-E at the Advanced Photon Source. This work was supported by the US National Institutes of Health (HL086392 to M.Z. and T32HL087745 to E.J.L.), the NYCOMPS that is supported by the NIH Protein Structure Initiatives PSI-II (GM075026 to W. A. Hendrickson), and the American Heart Association (0630148N to M.Z.). M.Z. is a Pew Scholar in Biomedical Sciences.

Author Contributions E.J.L. and M.Z. conceived and designed the experiments. E.J.L. purified and crystallized the protein; M.Q. performed and analysed the radiotracer flux and SPA binding assays; E.J.L. and M.Z. collected and processed the X-ray data, solved the structure, and wrote the paper.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession 3K3F and 3K3G. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.Z. (mz2140@columbia.edu).

doi:10.1038/nature08558 nature

METHODS

Homology screen, protein purification and crystallization. Human UT-B (NCBI accession NM_015865) was nominated as a target to NYCOMPS, and bacterial homologues were then selected ³⁶. A total of 14 UT genes were amplified by PCR from the genomic DNA of the following bacteria: Actinobacillus pleuropneumoniae, Bacteroides fragilis, Colwellia psychrerythraea, D. vulgaris, Nitrosomonas europaea, Ochrobactrum anthropi, Pseudomonas aeruginosa, Pseudomonas fluorescens, Pseudomonas putida, Staphylococcus epidermidis, Staphylococcus saprophyticus, Yersinia frederiksenii, Yersinia mollaretti and Yersinia pseudotuberculosis. Each gene was first cloned into modified pET plasmids (Invitrogen Inc.) that produced either an N- or a C-terminal His-tagged protein. Small-scale (21) test purifications were conducted, and D. vulgaris UT (dvUT) and Y. frederiksenii UT yielded stable detergent-solubilized protein as judged by elution as a single mono-dispersed peak from a size-exclusion column. The identity of the purified protein was verified by mass spectrometry. Only dvUT yielded diffracting crystals, and was the focus of further experiments.

To increase the expression level, dvUT was cloned into a modified pET-SUMO plasmid (Invitrogen) with an N-terminal polyhistidine tag and a SUMO domain. The protein was expressed in E. coli BL21(DE3) cells, solubilized with 30 mM DDM, and purified on a cobalt affinity column (Clontech Inc.). After cleavage of the His tag and SUMO domain by incubation with SUMO protease, the protein was exchanged into a buffer of 300 mM NaCl, 20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol and 40 mM low-purity *n*-octyl-β-Dmaltoside (Sol-Grade from Anatrace) on a Superdex 200 10/300 GL gel filtration column (GE Health Sciences). Each litre of cell culture yielded 0.2-0.3 mg of dvUT after the gel-filtration step. The final protein concentration was \sim 8 mg ml $^{-1}$ as approximated by ultraviolet absorbance. Native crystals were grown by vapour diffusion in unmixed sitting drops formed by combining 2 μl of the protein solution with an equal volume of well solution containing 22% polyethylene glycol PEG1500 and 100 mM sodium cacodylate, pH 6.5. Derivatized crystals were grown by including 10 mM thiomersal or potassium gold cyanide in the drop solution. Before flash-freezing in liquid nitrogen, the crystals were cryoprotected by gradually increasing the concentration of PEG1500 in the well solution to 45% over a period of 30 h. The native crystals diffracted to resolutions of up to 3.8 Å and indexed to the $P3_1$ space group, with unit cell dimensions a and c of 102.9 and 141.7 Å, respectively. Curiously, crystals grown from protein purified in high-purity OM (AnaGrade from Anatrace) uniformly failed to diffract to better than 4.5-5 Å. Inclusion of heavy atoms changed the space group to $P6_3$ with unit cell dimensions a and c equal to 110.13 and 84.86 Å, respectively, and improved resolution to up to 2.3 Å. The DMU-bound crystals were obtained by incubating the protein with 10 mM ligand at 20 °C for 30 min before setting up drops in the same condition used to obtain the gold-derivatized crystals.

Data collection and structure solution. Diffraction data were collected on beamlines X25 and X29 at the NSLS and on beamline 24ID-E at the Advanced Photon Source. Two data sets were used for structure solution: a 2.5 Å mercury data set collected at a wavelength of 1.01 Å, and a 2.3 Å gold data set collected at a wavelength of 1.04 Å. The data were indexed, integrated and scaled using the HKL2000 software suite³⁷. The mercury data set showed a stronger anomalous signal and was therefore used to obtain the initial phases. Eight heavy atom sites were located by SAD using the programs SHELX³⁸ and phenix.hyss³⁹. The programs Phaser⁴⁰ and RESOLVE as run by phenix.autosol were used to calculate experimental phases, carry out density modification, and build a partial model containing 213 of the total 337 residues. A further 32 residues were added manually in Coot⁴¹, and the improved model was used to calculate an improved map using combined experimental and model phases in PHASER. After density modification in DM⁴², ARP/wARP⁴³ was able to dock 189 residues in sequence. Iterative rounds of map calculation, density modification and manual building were used until the model was roughly 85% complete, and then it was used as a molecular replacement model for the gold data set with Molrep⁴⁴. The structure was completed using Refmac5 (ref. 45) and Coot. For the later cycles of refinement, four TLS groups identified by TLSMD⁴⁶ were included, and the model geometry was analysed using Molprobity. The final model was complete except for three C-terminal residues and four residues from a disordered loop. The final structure was then used as a model for molecular replacement for a 2.5 Å data set collected on the DMU-bound crystals. After rigid-body refinement of the molecular replacement solution, the DMU-bound model was refined by simulated annealing in CNS⁴⁷ and automatic and manual refinement in Refmac5 and Coot.

Tracer uptake in Xenopus laevis oocytes. dvUT and human UT-B (NCBI accession NM 015865) were cloned into a modified pBluescript vector for in vitro transcription, and the mRNAs were purified by the Trizol reagent (Invitrogen). For each measurement, ten oocytes were transferred to a transport vial that contained 360 µl transport buffer that completely covers the oocytes and is composed of 96 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂ and 5 mM HEPES, pH 7.6, and supplemented with 181 μ M of ¹⁴C-urea at 55 mCi mmol⁻¹. A timer was started immediately after oocyte transfer. The reaction was stopped at the desired time point by adding 4 ml of ice-cold buffer without radiotracer and aspirated after a brief mixture by swirling the vial. The oocytes were washed three times with the same buffer. After the last wash each oocyte was transferred into a separate scintillation vial that contained 200 µl of 10% SDS, and vortexed vigorously. After an oocyte was dissolved, 5 ml of scintillation cocktail was added to each vial and vortexed. Radioactivity in each vial was determined by liquid scintillation counting. For standards, $3 \times 5 \,\mu l$ of each transport buffer was used, plus 200 µl of 10% SDS and 5 ml scintillation cocktail.

SPA-based binding assay. Cu²⁺-coated YSi SPA beads (GE Healthcare) were diluted to 2.5 mg ml⁻¹ in 150 mM Tris/Mes, pH 7.5, with 50 mM NaCl, 20% glycerol, 1 mM TCEP (Sigma), 0.1% n-dodecyl-β-D-maltopyranoside (Anatrace Inc.) with purified His-tagged dvUT (50-250 ng per assay) and radiolabelled urea. One-hundred microlitres of SPA-bead/protein/radiotracer solution was added to individual wells of clear-bottom, white-wall 96-well plates. For isotopic dilution and competition experiments depicted in Fig. 1e the final concentration of ¹⁴C-urea was kept constant at 181 μM (55 mCi mmol⁻¹), whereas the concentration of non-labelled urea or DMU was increased from 0-50 mM. Saturation binding (as shown in Fig. 1d) was performed with increasing concentrations of ¹⁴C-urea at a specific activity of 2 mCi mmol⁻¹. Binding was performed in the dark for 30 min at 4 °C with vigorous shaking on a vibrating platform. To determine the non-specific background binding activity, 400 mM imidazole was added to the wells because imidazole competes with the His-tag of the recombinant protein. Plates were read in the SPA mode of a Wallac 1450 MicroBeta plate PMT counter. Total c.p.m. were corrected for non-specific binding by subtracting the c.p.m. of the samples in the presence of imidazole, yielding the specific c.p.m. To prevent radioligand depletion in the saturation binding experiments 250 ng of purified dvUT, a protein amount substantially below the capacity of the SPA beads, was used per assay. The transformation of c.p.m. to the amount of urea was calculated with a standard curve of known amounts of radioactivity (determined by adding scintillation cocktail to the samples), and the slope of this linear relation was used to transform c.p.m. into mol l-1 (ref. 17). All experiments were performed at least in duplicate with replicas of≥3 and data are expressed as mean ± standard error. Data fits of kinetic analyses were performed using nonlinear regression algorithms in Sigmaplot (SPSS Inc.), and error bars represent the s.e.m. of the fit.

- Punta, M. et al. Structural genomics target selection for the New York Consortium on Membrane Protein Structure. J. Struct. Funct. Genomics. doi:10.1007/s10969-009-9071-1 (in press).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
- Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX. J. Appl. Crystallogr. 37, 843–844 (2004).
- 39. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. Acta Crystallogr. D 55, 849–861 (1999).
- McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
- 41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta*
- Crystallogr. D 60, 2126–2132 (2004).
 42. Collaborative Computational Project number 4.. The CCP4 suite: programs for
- protein cystallography. Acta Crystallogr. D 50, 760–763 (1994).
 43. Potterton, L. et al. Developments in the CCP4 molecular-graphics project. Acta Crystallogr. D 60, 2288–2294 (2004).
- 44. Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. Model preparation in MOLREP and examples of model improvement using X-ray data. *Acta Crystallogr. D* **64**, 33–39 (2008).
- 45. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. D 60, 2184–2195 (2004).
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* 57, 122–133 (2001).
- Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D 54, 905–921 (1998).

ARTICLES

Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme

Yan Qi 1 , Marie C. Spong 2 , Kwangho Nam 2 , Anirban Banerjee 2 †, Sao Jiralerspong 2 †, Martin Karplus 2,6 & Gregory L. Verdine 1,2,3,4,5

How living systems detect the presence of genotoxic damage embedded in a million-fold excess of undamaged DNA is an unresolved question in biology. Here we have captured and structurally elucidated a base-excision DNA repair enzyme, MutM, at the stage of initial encounter with a damaged nucleobase, 8-oxoguanine (oxoG), nested within a DNA duplex. Three structures of intrahelical oxoG-encounter complexes are compared with sequence-matched structures containing a normal G base in place of an oxoG lesion. Although the protein-DNA interfaces in the matched complexes differ by only two atoms—those that distinguish oxoG from G—their pronounced structural differences indicate that MutM can detect a lesion in DNA even at the earliest stages of encounter. All-atom computer simulations show the pathway by which encounter of the enzyme with the lesion causes extrusion from the DNA duplex, and they elucidate the critical free energy difference between oxoG and G along the extrusion pathway.

Damaged nucleobases in DNA, generated spontaneously through the attack of reactive species on the genome, are an important source of carcinogenic mutations^{1,2}. Lesion-specific DNA glycosylases catalyse removal of most aberrant nucleobases and initiate the base-excision repair pathway^{3,4}. Searching through the vast genome to locate the exceedingly rare lesions (1 per 10⁶–10⁷ base pairs) without energy infusion represents an especially challenging task for DNA glycosylases, because the covalent structures of the target lesions often closely resemble their normal counterparts.

One lesion of particular interest is 8-oxoguanine (oxoG, Fig. 1a), an oxidation product of guanine⁵, which causes G•C to T•A transversion mutations during replication⁶. Despite the obvious structural similarity

of oxoG to the vastly more abundant G, and its rather innocuous effects on isolated DNA helices^{7–10}, oxoG is recognized specifically and efficiently by 8-oxoguanine DNA glycosylases, namely bacterial MutM and eukaryotic OGG1. High-resolution structures of complexes comprising MutM¹¹ and OGG1¹² bound to oxoG-containing DNA and poised for base excision (lesion-recognition complexes, or LRCs) have shown that both enzymes extrude oxoG completely from the DNA helical stack and insert the extrahelical lesion into a recognition pocket on the enzyme (Fig. 1b). Importantly, structures of LRCs depicting a diverse array of damaged nucleobases bound to DNA glycosylases have established that all enzymes acting on single-nucleobase lesions employ extrahelical base excision¹³. These findings have raised

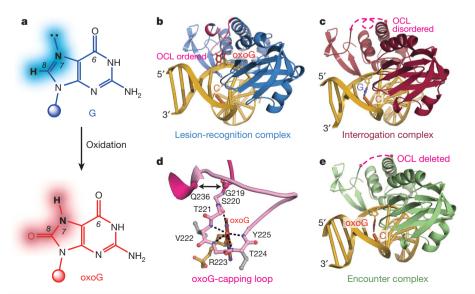


Figure 1 | Generation and recognition of 8-oxoguanine. a, Structural comparison of G versus oxoG, with differences highlighted. b, Overall structure of a lesion-recognition complex, LRC3, with an extrahelical oxoG lesion bound in the enzyme active site. c, Sequence-matched interrogation complex, IC3 (ref. 14), with a fully intrahelical target G•C base-pair. d, Close-up view of the oxoG-capping loop (OCL) and its contacts to oxoG in LRC3. The light pink region represents the residues deleted in EC3–5. Residues with grey side chains were mutated to prolines in the ECs bearing OCL point mutations. e, Sequence-matched encounter complex, EC3, bearing a fully intrahelical target oxoG•C base-pair.

¹Graduate Program in Biophysics, Harvard Medical School, Boston, Massachusetts 02115, USA. Departments of ²Chemistry and Chemical Biology, ³Molecular and Cellular Biology, and ⁴Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Chemical Biology Initiative and Program in Cancer Chemical Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁶Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université de Strasbourg, 8 allée Gasparol Monge, 67000 Strasbourg, France. †Present addresses: Rockefeller University, New York 10021, USA (A.B.); Department of Biochemistry and Molecular Biology, Division of Cancer Medicine, University of Texas, MD Anderson Centre, Houston, Texas 77030, USA (S.J.).

considerable interest in understanding the sequence of events that begins with a DNA glycosylase scanning DNA and encountering a lesion, and ends with insertion of the lesion into the active site followed by catalysis of base excision. Though much is known structurally about the end stage of this process, namely extrahelical lesion recognition and excision, virtually nothing is known about the nature of the initial encounter between the enzyme and an intrahelical lesion. Here we report the structures of three complexes representing the state of initial encounter between MutM and a fully intrahelical, base-paired oxoG lesion in DNA. The clearly discernible structural differences between the encounter complexes (ECs) versus sequence-matched complexes having a target G instead of oxoG (interrogation complexes, or ICs) indicate that MutM can discriminate an oxoG lesion from its normal counterpart even at the stage of intrahelical encounter. Computational simulations enabled by such closely matched structures indicate active participation of MutM in the early steps of lesion extrusion from DNA.

Entrapment strategy

To trap the ordinarily fleeting state of intrahelical lesion encounter, we introduced two key modifications: (1) disulphide-crosslinking¹⁴ to limit the roaming range of MutM on DNA, and (2) mutation of the oxoG-capping loop (OCL), which interacts extensively with the extrahelical oxoG (Fig. 1d), so as to prevent the LRC from dominating the thermodynamic landscape of MutM–DNA interactions. Biochemically, the MutM OCL variants failed to excise oxoG from DNA, but remained competent in cleaving DNA at abasic sites (Supplementary Fig. 2a–b). This activity profile indicates that OCL mutation only disrupts oxoG-recognition dependent reaction step(s) in the base excision cascade.

To isolate the effect of the lesion on DNA, we crystallized matched pairs of structures, with each pair differing only in whether the DNA component contains a G or an oxoG at the interrogation site by MutM. To establish the effect of sequence context on intrahelical lesion recognition, we structurally characterized matched EC/IC pairs and the corresponding LRCs with three different sequences (for example, EC3/IC3/LRC3, Supplementary Fig. 3).

Global structures of encounter complexes

The global structure of EC3 (1.89 Å resolution) closely resembles that of IC3 (Fig. 1c), with a sharp bend in the DNA localized to the binding site of MutM (Fig. 1e). The target oxoG•C base pair in EC3 is unambiguously intrahelical (Supplementary Fig. 4); EC3 is thus the first structure of a DNA glycosylase directly observed at the stage of its initial encounter with an intrahelical lesion. Importantly, EC3 yielded the same crystal form (Supplementary Table 1) that has captured diverse states of nucleobase extrusion by MutM, from unextruded to partially and fully extruded^{11,14}. Therefore, the unextruded recognition state in EC3 represents a preferred mode of MutM–DNA interaction when the fully extruded state is destabilized.

Three MutM residues serve critical roles in probing the target nucleobase. In LRC3, F114 and M77 intercalate into the DNA helix on the 3'-side of the broken oxoG•C pair, and R112 reaches into the

space formerly occupied by the oxoG to form a hydrogen bond with the opposite C (Fig. 2a). In the intrahelical state observed in EC3, only F114 penetrates the kinked DNA helix and buckles the target base-pair, whereas M77 is retracted. Most significantly, insertion of R112 into the helical stack is prevented by the intrahelical oxoG•C base-pair. R112 instead curls under to form a hydrogen bond with the N3-atom of the 3′-side G neighbouring the target C (Fig. 2c and Supplementary Fig. 5a), and all four nucleobases can interact with R112 in the equivalent minor groove position. As in IC3 (Fig. 2b), this alternative conformation of R112 is stabilized through hydrogen bonding with E78.

Despite the differences in DNA sequences and the crosslinking positions on the DNA, the structures of EC4 (1.62 Å resolution) and EC5 (1.83 Å) bear all defining global structural characteristics described above for EC3 (Supplementary Fig. 5b–c). The pronounced similarity among the EC structures indicates that the global features of the encounter between MutM and an intrahelical lesion are relatively independent of local sequence context.

MutM distinguishes a target oxoG from G

Each EC and its sequence-matched IC differ in the protein-DNA interface by only two atoms, those that distinguish oxoG from G (Fig. 1a). Despite several shared key features, each EC/IC pair bears significant elements of structural divergence in stacking and buckling in the DNA component, and in the protein-DNA interface (below and Supplementary Information). The most pronounced differences lie in the DNA backbone conformation. The target G in IC3 adopts the ground-state 2'-endo sugar pucker (Fig. 3a). Replacing G by oxoG results in pseudorotation of the sugar to an alternative pucker (C4'-exo) (Fig. 3c and Supplementary Fig. 9). An identical change in sugar pucker is observed in going from IC4 to EC4 (Supplementary Fig. 10a-c). This conformational reorganization of the DNA backbone is directly attributable to the 8-oxo substituent on oxoG, as modelling of an 8-oxo carbonyl onto the G in IC3 shows a steric clash with C2' (2.6 Å, Fig. 3b). Importantly, oxoG in isolated B-form DNA adopts the 2'-endo sugar pucker, with an 8-O/C2' distance (3.0 Å) beyond the repulsion range^{7,10}. This indicates that the aforementioned steric strain results from conformational manipulation by MutM while inspecting a target base in DNA. The mechanism of this effect is readily apparent from the crystal structures. Buckling of the target base pair by MutM (Supplementary Fig. 6) causes glycosidic bond rotation that thrusts C8 towards C2', making the C8/C2' distance shorter than in naked DNA. Whereas the C8-H in G is small enough to avoid a clash with C2', the much larger C8-O in oxoG jams into C2′, thereby introducing the clash that disfavours the 2'-endo sugar pucker. Interestingly, DNA polymerases employ a similar strategy to amplify the difference between oxoG-containing DNA and normal DNA15.

The IC5/EC5 pair also shows clear evidence of oxoG-specific intrahelical recognition, but of a different form from that in sets 3 and 4. In IC5, the target G already adopts the alternative C4'-exo sugar pucker (Fig. 3d), hence conversion to oxoG engenders no 8-oxo/2'-CH₂

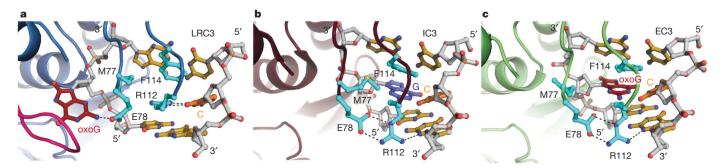


Figure 2 | Helix-penetration by MutM residues. a, LRC3; b, IC3; c, EC3. Colour-coding is as in Fig. 1, except side chains of the key residues (M77, R112, F114 and E78) are shown in cyan.

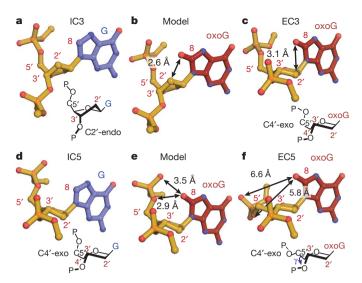


Figure 3 | OxoG-dependent local DNA structure alterations at the site of the target base in EC3 and EC5. $\bf a$ and $\bf d$, target $\bf G$; $\bf c$ and $\bf f$, target oxo $\bf G$; $\bf b$ and $\bf e$, models constructed by the addition of a carbonyl oxygen to the 8-position of the corresponding target $\bf G$ residue in $\bf a$ and $\bf d$, respectively. The perspectives were chosen to facilitate examination of sugar puckers (diagrammed below $\bf a$, $\bf c$, $\bf d$, and $\bf f$). Steric clashes in the models are indicated by double-headed arrows.

steric clash. However, the close proximity between the C8-H and 5′-phosphate of the target G in IC5 would result in unfavourable steric and electronic interactions between the 8-oxo carbonyl of an oxoG in the same conformation and its own 5′-phosphate (Fig. 3e). This clash is avoided in EC5 (Fig. 3f) by rotations about the C4′-C5′ bond and the phosphodiester linkage to the 5′-neighbouring nucleoside (Supplementary Fig. 10d). Again, the above DNA backbone adjustments are absent in isolated oxoG-containing DNA^{7,10}.

In conclusion, when MutM encounters an intrahelical oxoG, the protein can induce two distinct local modes of conformational change in the DNA backbone to avoid unfavourable interactions with the 8-oxo group: pseudorotation of the sugar as in EC3/4, and backbone rotations observed in all three ECs. The differences between EC3/4 and EC5 (Supplementary Figs 6, 7 and 11) might simply result from the non-identical constraints on helix architecture imposed by base-stacking with either C (EC3/4) or G (EC5) on the 5'-side of the lesion.

The structural changes on replacement of a normal target G by an oxoG lesion, although caused by local steric effects, are transmitted to the surrounding DNA backbone, causing consistent differences in the MutM–DNA interface across all three sets of structures. Most obviously, the DNA backbone around the lesion moves towards the LRC conformation in going from ICs to ECs (Supplementary Fig. 12); importantly, this takes place at positions distal from sites of crystal contacts (Supplementary Fig. 13). In general, more residues in the ECs directly contact the backbone of the lesion-containing strand than in the corresponding ICs, and the LRCs maintain these contacts (Supplementary Fig. 14). Although no single structural feature is absolutely diagnostic for the presence of an oxoG, an intrahelical lesion modifies the overall complex structure further towards the extruded

Energetics of lesion extrusion

To determine the thermodynamic and possible kinetic consequences of the structural perturbations described above, we performed molecular dynamics simulations to calculate the free energies for the extrusion of oxoG and G from DNA in the presence and absence of MutM. The availability of matched intrahelical and extrahelical structures (EC4/IC4/LRC4) enables computational studies of the nucleobase extrusion pathway, which consists of three successive events: (1) disruption of the target base pair and extrusion of the target base into the minor

groove; (2) rotation about the glycosidic bond from anti to syn torsion angle; and (3) swivelling of the target nucleoside away from the helix to an extended, fully extrahelical state (Fig. 4b and Supplementary Figs 15-17, Supplementary Movies 1 and 2). Although the simulations encompass all three stages of base extrusion, event (1) is of primary interest for understanding the origin of lesion recognition by MutM. Figure 4a plots free energies along the extrusion pathway from the initial intrahelical state to the lowest free-energy state in which the target base is extrahelical but not yet rotated and extended. In the presence of MutM, the overall barrier to oxoG extrusion is only \sim 4 kcal mol⁻¹ (Z = 0.45), whereas that for G is \sim 11 kcal mol⁻ (Z = 0.63, Supplementary Table 2). Rates for nucleobase extrusion were estimated from these free energy barriers based on earlier work¹⁶: $\sim 1 \times 10^{-3} \,\mathrm{ps}^{-1}$ for oxoG and $1 \times 10^{-8} \,\mathrm{ps}^{-1}$ for G (Supplementary Information). Simulations based on the EC5/IC5/LRC5 structures give similar trends (Supplementary Fig. 18, Supplementary Movies

What roles does MutM have in base extrusion? In the absence of MutM, bent DNA, assuming the conformation in EC4/IC4, experiences considerably reduced overall free energy barriers to target base extrusion, as compared to naked unbent DNA (Fig. 4a); the values obtained here for unbent DNA are close to those reported previously^{17–19}. Therefore, one mechanism by which MutM promotes nucleobase extrusion is by introducing a bend in the DNA near the enzyme active site, as the loss of stacking interactions for the target base²⁰ destabilizes the intrahelical state. Furthermore, comparison of the free energy barriers for the naked bent DNA with those for MutM-bound DNA indicates that MutM actively facilitates the target nucleobase extrusion, and the facilitation effect is greater for oxoG than G. This additional reduction of the free energy for base extrusion involves several MutM residues (Fig. 5 and Supplementary Fig. 15). F114 and M77 buckle the target base pair to interrupt base-stacking interactions. Most striking are the effects of R112. At the initial MutM-DNA encounter stage, R112 remains close to the non-Watson-Crick (WC) lone pair on O2 of the C opposite the target base, and this serves as the wedge point for helix invasion (Fig. 5a). R112 moves from its interaction with the non-WC to the WC lone pair of O2-C, in direct competition with the target nucleobase (Fig. 5b, c). From this 'anchored' position, R112 invades the helix to establish bidentate hydrogen bonds with O2 and N3 of C, concomitant with breakage of the target base-pair (Fig. 5d-f). To confirm the role of R112 in promoting target base-pair disruption, free energy simulations for MutM-bound DNA were repeated with the interactions between the R112 side chain and the target oxoG•C or G•C turned off. R112 now fails to invade the DNA helix, the disruption of the target base pair is delayed, and both oxoG and G experience elevated free energies at Z = 1.0 (Supplementary Fig. 19). A similar arginine residue (R128) in O⁶-alkylguanine-DNA alkyltransferase has been observed computationally to invade the DNA helix and disrupt the target base-pair following spontaneous base-pair fluctuations¹⁶.

How does MutM locate oxoG in the genome?

Recent single-molecule tracking studies have established that MutM translocates along DNA with a nearly negligible free energy barrier (~2 kcal mol⁻¹ per base pair, ref. 21). This effectively rules out the possibility that MutM must inspect every nucleobase in its extrahelical lesion recognition pocket to locate oxoG. MutM must therefore possess a mechanism for discriminating intrahelical lesions from normal bases. The absence of a thermodynamic preference for the MutM OCL variants to bind oxoG•C-containing DNA (<twofold, Supplementary Fig. 2c) indicates that intrahelical lesion recognition by MutM must have a kinetic origin. This conclusion is supported by biochemical experiments on wild-type MutM²²⁻²⁴. The matched sets of structures reported here indicate that encounter of a lesion by MutM promotes an 'extrudogenic' DNA backbone conformation that is conducive to nucleobase extrusion. Molecular dynamics simulations of the base extrusion pathway based on these structures clearly show that, in the presence of

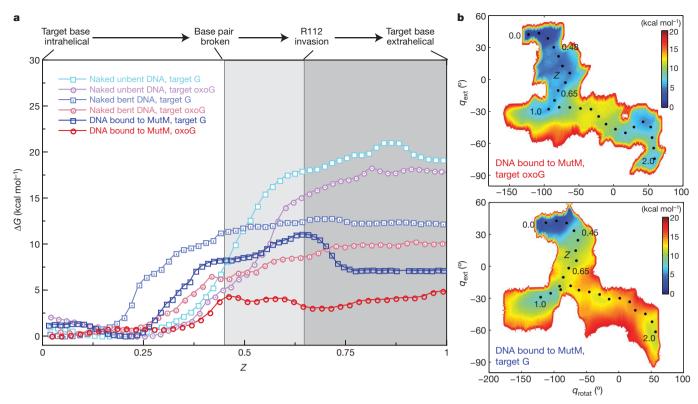


Figure 4 | Free energy profiles of nucleobase extrusion imply active participation of MutM. a, Free energies along the base extrusion pathways, as a function of the normalized arc length, Z, in b and Supplementary Fig. 17. Critical events associated with main energetic barriers are marked. b, Free energy landscape plots for extrusion of oxoG starting from EC4 (top) and G

from IC4 (bottom). See Supplementary Fig. 16 for definitions of $q_{\rm ext}$ and $q_{\rm rotat}$. The black dotted lines denote the minimal free energy paths for base extrusion, which define Z. The lowest free energy points in the intrahelical states are set to zero.

MutM, an oxoG lesion encounters a much lower and flatter energy barrier for initial extrusion from the helical stack than does G. In both ECs and ICs, a cherry-picker residue in MutM, R112, promotes the disruption of the target pair by competing with Watson–Crick base-pairing. This active intrahelical interrogation and extrusion mechanism differs from the passive mechanism proposed for uracil DNA glycosylase (UNG), in which the protein is envisaged to preferentially capture a spontaneously extruded lesion²⁵.

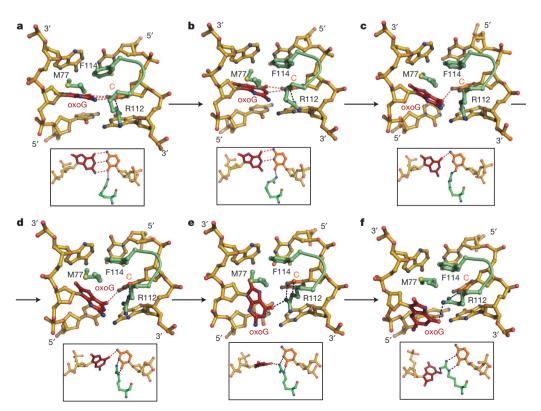


Figure 5 | R112-catalysed oxoG extrusion. a-f, Snapshots from targeted molecular dynamics (TMD) simulations in the vicinity of the target base, showing helix invasion by R112 and concomitant extrusion of the target oxoG. The three DNA helix-invading MutM residues (M77, R112 and F114) are coloured in green. Nucleosides around the target oxoG•C pair are shown in gold. The inset figure of each panel is an orthogonal view. See the text for the progression of events **a-f**. Refer to the Supplementary Movies for the complete base extrusion trajectories of oxoG and G by MutM.

The massively redundant Brownian search mechanism employed by MutM to inspect DNA requires minimization of the kinetically expensive extrusion of normal bases. MutM initially encounters a fully intrahelical, base-paired oxoG lesion. Through a combination of DNA bending, unstacking, buckling of the target pair and helix invasion by R112, MutM can discriminate a lesion from an undamaged base and preferentially promotes oxoG extrusion. This mechanism indicated by the structures and computational studies here provides clues for the kinetic efficiency of MutM in lesion search and recognition.

METHODS SUMMARY

Geobacillus stearothermophilus MutM with crosslinking mutation Q166C and OCL mutation (Δ220–235, V222P or T224P) was overexpressed in Escherichia coli and purified as described before¹¹. DNA oligomers were synthesized with standard phosphoramidite chemistry and the disulphide crosslinker was introduced using H-phosphonate chemistry¹⁴. Purified crosslinker-containing DNA duplex (10 μM) was mixed with 20 μM MutM in 20 mM Tris pH 7.4 and 50 mM NaCl for 2–3 days. The crosslinked complex was purified from the mixture with a Mono Q column, concentrated to 225–250 μM in 50 mM NaCl, 20 mM Tris pH 7.4, and crystallized in reservoir solutions containing 12–18% PEG 8K, 100 mM sodium cacodylate pH 7.0 and 5% glycerol. Diffraction data were collected on synchrotron light sources with cryo-protected crystals. Structures were solved by molecular replacement using coordinates of the protein part from isomorphous structures of MutM crosslinked to undamaged DNA.

On the basis of the EC4/IC4/LRC4 crystal structures, four end-state systems with MutM and oxoG-containing or normal DNA were prepared, energy minimized, and equilibrated using molecular dynamics simulations. The all-atom CHARMM force fields^{26–28} and TIP3P water model²⁹ were used. All simulations were performed with orthorhombic periodic boundary conditions, and electrostatic interactions were evaluated using the particle mesh Ewald summation method³⁰. The base extrusion pathways for oxoG and G were determined by employing steered and targeted molecular dynamics simulations^{31–33}, and the free energies were evaluated using umbrella sampling free energy simulations³⁴. Similar procedures were applied for simulations with naked bent and unbent DNA duplexes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 May; accepted 7 October 2009.

- Lindahl, T. Instability and decay of the primary structure of DNA. Nature 362, 709–715 (1993).
- 2. Loeb, L. A. A mutator phenotype in cancer. Cancer Res. 61, 3230–3239 (2001).
- Barnes, D. E. & Lindahl, T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.* 38, 445–476 (2004).
- Fromme, J. C. & Verdine, G. L. Base excision repair. Adv. Protein Chem. 69, 1–41 (2004).
- Grollman, A. P. & Moriya, M. Mutagenesis by 8-oxoguanine: an enemy within. Trends Genet. 9, 246–249 (1993).
- Shibutani, S., Takeshita, M. & Grollman, A. P. Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* 349, 431–434 (1991).
- Lipscomb, L. A. et al. X-ray structure of a DNA decamer containing 7,8-dihydro-8oxoguanine. Proc. Natl Acad. Sci. USA 92, 719–723 (1995).
- Oda, Y. et al. NMR studies of a DNA containing 8-hydroxydeoxyguanosine. Nucleic Acids Res. 19, 1407–1412 (1991).
- Plum, G. E., Grollman, A. P., Johnson, F. & Breslauer, K. J. Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex. *Biochemistry* 34, 16148–16160 (1995).
- Bowman, B. R., Lee, S., Wang, S. & Verdine, G. L. Structure of the E. coli DNA glycosylase AlkA bound to the ends of duplex DNA: a system for the structure determination of lesion-containing DNA. Structure 16, 1166–1174 (2008).
- Fromme, J. C. & Verdine, G. L. DNA lesion recognition by the bacterial repair enzyme MutM. J. Biol. Chem. 278, 51543–51548 (2003).
- Bruner, S. D., Norman, D. P. G. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* 403, 859–866 (2000).
- 13. Zharkov, D. O. Base excision DNA repair. *Cell. Mol. Life Sci.* **65**, 1544–1565 (2008).
- Banerjee, A., Santos, W. L. & Verdine, G. L. Structure of a DNA glycosylase searching for lesions. Science 311, 1153–1157 (2006).
- Hsu, G. W., Ober, M., Carell, T. & Beese, L. S. Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature* 431, 217–221 (2004).

- Hu, J., Ma, A. & Dinner, A. R. A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl Acad. Sci. USA* 105, 4615–4620 (2008).
- 17. Priyakumar, U. D. & Mackerell, A. D. Jr. NMR imino proton exchange experiments on duplex DNA primarily monitor the opening of purine bases. *J. Am. Chem. Soc.* 128, 678–679 (2006).
- Banavali, N. K. & MacKerell, A. D. Jr. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. J. Mol. Biol. 319, 141–160 (2002).
- Cheng, X. et al. Dynamic behavior of DNA base pairs containing 8-oxoguanine. J. Am. Chem. Soc. 127, 13906–13918 (2005).
- 20. Yang, W. Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst.)* 5, 654–666 (2006).
- Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA* 103, 5752–5757 (2006).
- Minetti, C. A. et al. Energetics of lesion recognition by a DNA repair protein: thermodynamic characterization of formamidopyrimidine-glycosylase (Fpg) interactions with damaged DNA duplexes. J. Mol. Biol. 328, 1047–1060 (2003).
- 23. Fedorova, O. S. *et al.* Stopped-flow kinetic studies of the interaction between *Escherichia coli* Fpg protein and DNA substrates. *Biochemistry* **41**, 1520–1528 (2002).
- Ishchenko, A. A. et al. Thermodynamic, kinetic, and structural basis for recognition and repair of 8-oxoguanine in DNA by Fpg protein from Escherichia coli. Biochemistry 41, 7540-7548 (2002).
- Parker, J. B. et al. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. Nature 449, 433–437 (2007).
- 26. MacKerell, A. D. Jr et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616 (1998).
- Mackerell, A. D. Jr & Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. J. Comput. Chem. 21, 105–120 (2000).
- Foloppe, N. & Mackerell, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* 21, 86–104 (2000).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935 (1983).
- 30. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089 (1993).
- van der Vaart, A. & Karplus, M. Simulation of conformational transitions by the restricted perturbation-targeted molecular dynamics method. J. Chem. Phys. 122, 114903 (2005).
- 32. Paci, E. & Karplus, M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Biol.* **288**, 441–459 (1999).
- 33. Hu, J., Ma, A. & Dinner, A. R. Bias annealing: A method for obtaining transition paths *de novo. J. Chem. Phys.* **125**, 114101 (2006).
- 34. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23**, 187 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the NIH: GM044853, GM047467, CA100742 (G.L.V.) and GM030804 (M.K.). Y.Q. is supported by a predoctoral fellowship from the National Science Foundation; M.C.S. by a predoctoral fellowship from the Howard Hughes Medical Institute; K.N. by a postdoctoral fellowship from National Cancer Center. We thank the staff of the NSLS, APS and CHESS synchrotron facilities. These experiments made use of the computing facilities at NERSC and Harvard FAS. We are grateful to J. Jin for experimental assistance and J. Pu, V. Ovchinnikov and other members of the Karplus and Verdine groups for helpful advice.

Author Contributions Y.Q., M.C.S., K.N. and A.B. contributed equally to the study. A.B., S.J. and M.C.S. cloned the constructs. A.B., M.C.S. and S.J. performed the biochemical and FP assays. A.B., S.J., M.C.S. and Y.Q. purified, crystallized and collected X-ray diffraction data and solved structures (A.B., S.J.: EC4, EC5; M.C.S.: EC3, EC3_{Y222P}, EC3_{T224P}, IC4, LRC5; Y.Q.: LRC3, EC3_{T224P}, EC5, IC5, LRC5). A.B. and G.L.V. designed the trapping strategy and crystallographic studies. K.N. and M.K. designed the computational studies, which K.N. then performed. A.B., M.C.S., Y.Q., K.N., G.L.V. and M.K. analysed data and wrote the paper. G.L.V. and M.K. directed the research. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3GPY (LRC3), 3GO8 (EC3), 3GP1 (EC3_{V222P}), 3GPP (EC3_{T224P}), 3GPU (EC4), 3GPX (IC4), 3GQ4 (LRC5), 3GQ3(EC5) and 3GQ5 (IC5). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.K. (marci@tammy.harvard.edu) or G.L.V. (gregory_verdine@harvard.edu).

doi:10.1038/nature08561 nature

METHODS

Cloning, overexpression and purification of wild-type and mutant MutM. Point mutants (V222P and T224P) and a deletion construct (Δ 220–235) of *Geobacillus stearothermophilus* MutM were introduced to the parent Q166C MutM construct¹⁴ in the pET24 (Novagen) expression vector using a QuikChange II site-directed mutagenesis kit (Stratagene). Wild-type (Q166C) and mutant proteins were overexpressed in *E. coli* and purified essentially as described before¹¹.

DNA synthesis and purification. Unless stated otherwise, all DNA oligomers were synthesized on an ABI 392 DNA synthesizer using standard reagents, purified by urea-PAGE, desalted with Sep-pak columns (Waters) and annealed with the complementary strands in 10 mM Tris pH 8.0. Crosslinker-containing DNA oligomers were synthesized as described before¹⁴. Briefly, a disulphide bearing tether was introduced at the non-bridging position of a nucleotide using H-phosphonate chemistry. The sequences of the DNA duplexes used for the crystallography studies are shown in Supplementary Fig. 3.

Preparative crosslinking reactions. ECs and IC4 were made by crosslinking the corresponding DNA oligomers to the Q166C MutM OCL deletion protein, whereas EC3 $_{\rm V222P}$ and IC5 contained Q166C V222P MutM and EC3 $_{\rm T224P}$ contained Q166C T224P MutM. LRCs were made using an enzymatically inactive mutant E3Q Q166C MutM. Preparative crosslinking reactions were performed by mixing 10 μ M crosslinker-containing DNA duplex with 20 μ M protein in 20 mM Tris pH 7.4 and 50 mM NaCl for 2–3 days. Purification of a crosslinked complex was performed on a MonoQ column using a linear salt gradient of 90% buffer A (20 mM Tris pH 7.4) to 60% buffer B (1 M NaCl, 20 mM Tris pH 7.4). Fractions containing the complex were pooled, concentrated and buffer exchanged into 50 mM NaCl, 20 mM Tris pH 7.4 and concentrated typically to 225–250 μ M before setting up crystallization experiments.

Crystallization, data collection and structure determination. The crosslinked complexes were crystallized in reservoir solutions containing 12–18% PEG 8K, 100 mM sodium cacodylate pH 7.0 and 5% glycerol. Microseeding with similar complex crystals was used for LRC3 and IC5. Crystals were allowed to grow to full size in about 3 weeks and were briefly soaked in a cryoprotectant solution containing 18% PEG 8K, 100 mM sodium cacodylate pH 7.0 and 25% glycerol before flash freezing in liquid nitrogen. Diffraction data was collected at A1 beamline of CHESS (Cornell High-Energy Synchrotron Source), X25 and X29 beamlines of the National Synchrotron Light Source, and 29ID-C and 29ID-E beamlines of the Argonne Photon Source. Data was processed using the HKL2000 program suites³⁵.

The coordinates of the protein from the isomorphous structure of MutM crosslinked to undamaged DNA14 were used as the initial model for refinement in CNS³⁶. The residues involved in protein–DNA interaction were omitted from the initial model. A rigid body fit followed by energy minimization and simulated annealing resulted in a partial model. Protein side chains were manually readjusted in COOT³⁷ followed by simulated annealing and grouped B factor refinement. The DNA was built into the readily discernible $F_0 - F_c$ map, and the density for the crosslink served as a useful register. Simulated-annealing omit maps were frequently used to reduce model bias. After all the visible protein and DNA atoms were built and $R_{\rm free}$ fell below 28%, individual B factor refinement was introduced. Water molecules were added to the model using both automated methods (in CNS) and manual inspection of difference maps. A final round of TLS refinements was done using Refmac5 (ref. 38). The TLS groups for the protein were defined using TLSDM³⁹ and the two DNA strands served as a single TLS group. Electron density for the oxoG-capping loops in ECs and ICs was not evident and these residues were omitted from the models. The disulphide tether was not built in any of the models. Amino acid side-chains of some residues were truncated at the α -, β -, γ -, or δ -carbon positions if electron density was not visible for the full side-chain.

Cleavage assay. DNA oligomer 5'-TGCGTCC°GAGTCTACC-3' was 5'-end labelled using T4 polynucleotide kinase (New England Biolabs) and $[\gamma^{-32}P]$ ATP (Perkin Elmer). Radiolabelled oxoG oligomer was annealed with 1.1-fold excess unlabelled complementary oligomer 5'-AGGTAGACTCGG ACGC-3'. Single turnover cleavage assays were carried out with 100 nM duplex DNA and 500 nM protein in a standard reaction buffer of 40 mM NaCl, 50 mM Tris pH 7.4 at 25 °C. For cleavage assays where abasic DNA was used, 5'-[32P]labelled uracil containing 16-mer duplex DNA (5'-AGCGTCCAUGTCTACC-3' purchased from Operon, and annealed to 5'-AGGTAGACTCGGACGC-3') was pretreated with uracil N-glycosylase (New England Biolabs) at 37 °C for 1 h. The resulting abasic site containing duplex DNA was used in the cleavage assays. Duplex DNA (100 nM) was incubated with 100 nM protein in a reaction buffer of 200 mM NaCl, 50 mM Tris pH 7.4 at 25 °C. Aliquots for both cleavage assays were taken out periodically and quenched by adding to equal volumes of 95% formamide in 1× TBE. Extent of cleavage was analysed by denaturing urea-PAGE and visualization on a storage phosphorimaging plate.

Binding assay. DNA oligomer labelled on the 3'-end with a fluorescein modification, 5'-TGGTAGACCTGGACGCF-3' (Operon, F indicates the fluorescein modification) was annealed to a complementary oligomer, 5'-AGCGTCCA°G GTCTACC-3', or 5'-AGCGTCCAGGTCTACC-3'. MutM–DNA binding measurements were carried out at 50 mM NaCl, 20 mM Tris pH 7.4 and 0.25 mM BME. Protein and DNA (10 nM) were mixed at different ratios in 96-well plates and equilibrated at room temperature. Changes in fluorescence polarization were monitored using excitation at 485 nm and emission at 530 nm on a SpectraMax M5 microplate reader (Molecular Devices), and data points were measured in triplicate. Dissociation constants were estimated by plotting the change in fluorescence polarization versus log of protein concentration using Kaleidagraph 3.6.

System setup for base extrusion simulations. The sequence-matched crystal structures of EC4 and LRC4 with an oxoG-containing DNA duplex and of IC4 with G at the site of interrogated oxoG were used for the simulations shown in Fig. 4 and Supplementary Figs 15, 17 and 19. Four systems were studied: (1) oxoG at the intrahelical interrogation site of MutM based on the EC4 structure; (2) G at the intrahelical interrogation site of MutM based on the IC4 structure; (3) oxoG in the extrahelical active site of MutM based on the LRC4 structure; and (4) G in the extrahelical active site of MutM with oxoG changed to G in the LRC4 structure by replacing the O8 atom from oxoG with an H, removing the H from N7, and introducing all the force fields parameters for G. The disordered oxoGcapping loops in EC4 and IC4 were built based on the ordered loop structure in LRC4 (residues 215–239). The E3Q mutation in LRC4 was modelled back to the wild-type E3 residue. The positions of hydrogen atoms were determined using the HBUILD facility in the program CHARMM40. Protonation states for all ionizable residues were chosen corresponding to pH 7, except that protonation states of histidine residues (on N_{ϵ} and N_{δ} atoms) were determined on the basis of possible hydrogen-bond interactions deduced from the crystal structures; the cysteine residues of the zinc fingers were deprotonated to maintain stable interactions with Zn²⁺ ion. For the DNA duplex, we kept the 11 central nucleotide base pairs surrounding the target base, and discarded the remaining part to retain the same number of nucleotides in the intrahelical and extrahelical conformations. The total charge of the resulting protein and DNA duplex was -13 e, which was neutralized by placing 13 Na⁺ ions 4.5 Å away from the bisectors of phosphate groups. All crystal waters were included. The resulting systems were further solvated with a box of water molecules of $75 \times 60 \times 60 \text{ Å}^3$, and water molecules within 2.5 Å of any non-hydrogen atoms of the protein, nucleic acid, Na⁺ ions and crystal waters were removed. The same procedure was repeated in the setup of other MutM-bound DNA complex systems. The final model of LRC4 containing an oxoG in the active site, for example, has 26088 atoms: 5024 protein and DNA atoms, 13 Na⁺ ions, and 7017 water molecules. Four systems with MutM bound to the DNA were prepared in a corresponding manner for the simulations starting with the EC5, IC5 and LRC5 structures shown in Supplementary Fig. 18.

For the naked bent DNA duplex simulations, the starting structures of the intrahelical and extrahelical oxoG conformations with the protein removed were taken from EC4 and LRC4, respectively. The systems were neutralized with 20 Na^+ ions and solvated with a box of water molecules of $50 \times 40 \times 40 \text{ Å}^3$. Again, water molecules within 2.5 Å from any non-hydrogen atoms of DNA and ions were removed. The naked unbent DNA duplex with the sequence of EC4 was model-built as a standard B-form DNA using Nucleic Acid Builder⁴¹ and solvated by applying the same procedure as for bent DNA, resulting in 697 DNA atoms, 20 Na⁺ ions and 2260 water molecules. All systems were duplicated with G at the site of oxoG to generate the corresponding intrahelical and extrahelical G conformations. Periodic boundary conditions were used for the simulations. Energy minimizations and molecular dynamics simulations. To alleviate high energy contacts of hydrogen atoms with other atoms, a total of 5000 steps of energy minimizations were performed for each system with a series of restraints by using conjugate gradient (CONJ) energy minimizations and Adopted Basis Newton-Raphson (ABNR) energy minimizations⁴⁰. Following the restrained energy minimizations to remove bad contacts, each system was equilibrated for a total of 1.7 ns. Details of energy minimizations and equilibrations are provided in Supplementary Information. The leapfrog Verlet algorithm⁴² was used with a 2 fs integration time step and SHAKE⁴³ was applied to bonds involving hydrogens. The final temperature was maintained at 298 K by coupling to an external thermal bath44, and unless otherwise noted, the volume was held constant. For the electrostatics, the particle mesh Ewald summation (PME) method³⁰ was used throughout: the calculations used the Ewald κ value of $0.340\, \mathring{A}^{-1}$ and the approximate grid spacing of $1.0\,\mathring{A}~(80\times 60\times 60$ fast Fourier transform grid). The real space terms in the PME method and the van der Waals interactions were evaluated with a cutoff distance of 9.0 Å, in which the latter interactions were smoothly turned off at the cutoff distance. The nonbonded interaction pair list and image list were updated at every 25 MD step doi:10.1038/nature08561 nature

and 100 MD step, respectively. The protein, nucleic acids, and ions are represented with the all-atom CHARMM 27 force fields^{26–28}, and the TIP3P model²⁹ was used for water molecules. For the 8-oxoguanine base, force field parameters were determined following the procedure suggested previously²⁸ and provided in Supplementary Table 3. All calculations were performed with the CHARMM program⁴⁰ (version c33a2).

Targeted molecular dynamics (TMD) simulations. To generate the base extrusion pathway for oxoG and G, TMD simulations $^{31-33,45}$ were carried out in the direction from the extrahelical active site bound configuration to the intrahelical configuration for the MutM-bound DNA complex, bent DNA, and unbent DNA duplex, respectively. Base extrusion and rotation were simulated by employing what is often called a steered molecular dynamics (SMD) algorithm; it is essentially the approach used previously 32,33 . The steering potential, $U_{\rm SMD}(t)$ at time t is defined as

$$U_{\text{SMD}}(t) = k(q_s(t) - q_0(t))^2 H(q_s(t) - q_0(t))$$

where k is a force constant, q_s is an order parameter for the base extrusion process, q_0 is the reference value of q_s at a given time, and H(x) is a step function that equals 1 when x is positive and is 0 otherwise, that is, a half harmonic potential is used to bias the system. The order parameter q_s was defined as the difference of two pseudo-dihedral angles,

$$q_{\rm s} = q_{\rm rotat} - q_{\rm ext}$$

where $q_{\rm rotat}$ and $q_{\rm ext}$ are the base rotation and base extrusion pseudo-dihedral angle⁴⁶ defined in Supplementary Fig. 16, respectively. The order parameter $q_{\rm s}$ initially had a positive value when the target base was in its extrahelical configuration, and decreased to negative values as the base re-anneals into the intrahelical configuration. During the simulation, when $q_{\rm s}$ was closer to the target value than q_0 , the dynamics were unperturbed, and q_0 was advanced towards the final target value by a preset amount ($\Delta q_{\rm s}$).

As described in the text, there is an essential structural rearrangement at the protein-DNA interface, as the base is extruded from the intrahelical conformation to the extrahelical conformation, that is, the invasion of R112, the formation of the ordered OCL, and importantly, the more global movement of the DNA backbone around the lesion towards the protein. The change in the enzyme-DNA interface that occurs during the base extrusion process was introduced by employing the restricted-perturbation targeted molecular dynamics (RP-TMD) algorithm31. In Supplementary Fig. 20, we represent the DNA-protein interface atoms to which the RP-TMD perturbation was applied. Several parameters have to be selected to perform the SMD and RP-TMD simulations: k and Δq_s for SMD and the maximum allowed perturbation in RP-TMD (see Supplementary Information). The selected parameters for SMD are 0.025 kcal mol⁻¹ degree⁻² for k and 0.025 degree for Δq_s . For the RP-TMD method, the total positional perturbation in all the atoms was restricted to 0.01 Å; if the RP-TMD perturbation was in the opposite direction to the force owing to the CHARMM potential, the positional perturbation was restricted to be 0.005 Å. During TMD simulations, we saved coordinates and velocities every 10 ps for the umbrella sampling free energy simulations described below.

Free energy simulations. To determine the free energy of base extrusion and insertion into the active site of MutM, two-dimensional umbrella sampling MD simulations were performed along the base extrusion pathway (generated from TMD simulations) with $q_{\rm ext}$ and $q_{\rm rotat}$ (Supplementary Fig. 16) as the reaction coordinates. Each umbrella sampling window was separated by 5° from neighbouring simulation windows and was run for 50 ps with a force constant of 0.05 kcal mol $^{-1}$ degree for each reaction coordinate: the total number of windows was 591 for MutM-bound DNA with oxoG, 656 for MutM-bound DNA with G, 511 for bent DNA with oxoG, 649 for bent DNA with G, 625 for unbent DNA with oxoG, and 664 for unbent DNA with G, respectively. The potentials of mean force (pmf) were computed using the weighted histogram

analysis method^{47,48}. For the naked bent DNA, we maintained the DNA bending by harmonically restraining certain sugar-phosphate backbone atoms to each saved TMD coordinate. The restrained atoms included the sugar-phosphate backbone atoms that were not terminal nucleotides and that were more than two nucleotides away from the target base-pair and the harmonic force constant was one tenth of the molar mass of each atom (kcal $\text{mol}^{-1} \mathring{A}^{-2}$). The convergence of computed free energies was tested by carrying out additional 50 ps umbrella sampling free energy simulations for the oxoG extrusion in the MutM-bound DNA complex. In addition, we repeated the TMD and free energy simulations for the EC5/IC5/LRC5 set (set 5) to confirm that the simulation results are not dependent on the local sequence context. The results for the oxoG and G simulations are presented in Supplementary Fig. 18 and Supplementary Movies. The initial free energy barriers for set 5 are similar to those of set 4, also the computed free energy barrier for the extrusion of oxoG is consistently lower than that for G. In set 5, the free energies of the extrahelical intermediate states (Z=1.0) are significantly lower than the values for set 4. This difference is attributed to the destabilization of the intrahelical conformation of set 5 owing to the weakened base stacking interactions with the 5'-side base. Because the extrahelical intermediate states are almost identical for set 4 and 5, the intrahelical free energy difference 'tilts' the free energy diagram in the observed way. To address the effects of R112 on the base extrusion shown in Supplementary Fig. 19, the nonbonded interactions between the side chain of R112 and the target oxoG•C or G•C base pairs were set to zero and free energy simulations were performed to obtain the pmf results; all other interactions of R112 were

- 35. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. et al. Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. Acta Crystallogr. D 54, 905–921 (1998).
- Emsley, P. & Cowan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* 57, 122–133 (2001).
- Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* 62, 439–450 (2006).
- Brooks, B. R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4, 187 (1983).
- 41. Macke, T. & Case, D. A. in *Molecular modeling of nucleic acids* (eds Leontes, N. B. & SantaLucia, J. Jr) 379–393 (American Chemical Society, 1998).
- Allen, M. P. & Tildesley, D. J. Computer simulation of liquids (Oxford Univ. Press, 1989).
- 43. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
- Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684 (1984).
- 45. Schlitter, J., Engels, M., Krueger, P., Jacoby, E. & Wollmer, A. Targeted molecular dynamics simulation of conformational change-application to the T↔R transition in insulin. *Mol. Simul.* 10, 291–308 (1993).
- 46. Banavali, N. K. & MacKerell, A. D. Jr. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. J. Mol. Biol. 319, 141–160 (2002).
- 47. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021 (1992).
- Rajamani, R., Naidoo, K. J. & Gao, J. Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution. J. Comput. Chem. 24, 1775–1781 (2003).

LETTERS

Ten per cent polarized optical emission from GRB 090102

I. A. Steele¹, C. G. Mundell¹, R. J. Smith¹, S. Kobayashi¹ & C. Guidorzi²

The nature of the jets and the role of magnetic fields in gamma-ray bursts (GRBs) remains unclear^{1,2}. In a baryon-dominated jet only weak, tangled fields generated *in situ* through shocks would be present³. In an alternative model, jets are threaded with large-scale magnetic fields that originate at the central engine and that accelerate and collimate the material⁴. To distinguish between the models the degree of polarization in early-time emission must be measured; however, previous claims of gamma-ray polarization have been controversial⁵⁻⁸. Here we report that the early optical emission from GRB 090102 was polarized at 10 ± 1 per cent, indicating the presence of large-scale fields originating in the expanding fireball. If the degree of polarization and its position angle were variable on timescales shorter than our 60-second exposure, then the peak polarization may have been larger than ten per cent.

The standard GRB fireball model³ comprises an initial compact emitting region, expanding relativistically, in which internal shocks dissipate the bulk energy, converting kinetic to radiated energy, the so-called prompt emission. As the shell of the relativistically expanding fireball collides with the surrounding circumburst medium, a forward shock is produced, which propagates outwards through the external medium and results in a long-lived afterglow. The afterglow's emission is detectable from X-ray to optical, infrared and, in some cases, radio wavelengths. Interaction of the relativistic fireball with the ambient medium also produces a short-lived reverse shock that propagates backwards through the expanding shell^{9,10}.

Exploiting the ability of robotic optical telescopes to respond rapidly and automatically to the discovery of new GRBs, a custom, fast-response, optical polarimeter¹¹ (RINGO) was deployed on the 2.0-m robotic Liverpool Telescope¹² (La Palma) with the goal of measuring the degree of polarization of optical emission from GRBs at early times. RINGO uses a rotating Polaroid to modulate the incoming beam, followed by corotating deviating optics that transfer each star image into a ring that is recorded on a chargecoupled device (CCD) (Fig. 1). Any polarization signal present in the incoming light is mapped out around the ring in a $\sin 2\theta$ pattern. RINGO was first used in 2006, when it observed GRB 060418 at 203 s after the GRB, coincident with the time of deceleration of the fireball. At this time the reverse-shock (assuming it was present) and forwardshock components would have contributed equally to the observed light. For GRB 060418 a 2σ upper limit on optical polarization of P < 8% was measured in the combined light from the emitting regions¹³. Until the burst reported here this was the only limit on early-time optical GRB polarization.

GRB 090102 was detected by the Swift satellite on 2 January 2009 at 02:55:45 Universal Time (UT), with a pulse of gamma rays lasting $T_{90} = 27$ s and comprising four overlapping peaks starting 14 s before the trigger time¹⁴. The automatic localization provided by the spacecraft was communicated to ground-based facilities, and a single 60-s RINGO exposure was obtained starting 160.8 s after the trigger time.

Simultaneously with our polarization observation of GRB 090102, a number of automated photometric follow-ups were also performed by other facilities 14-16. The optical light curve, beginning at 40-s postburst, can be fitted by a broken power law whose flux density F decays as a function of time $t(F \propto t^{-\alpha})$ with a gradient $\alpha = 1.50 \pm 0.06$ that then flattens to $\alpha = 0.97 \pm 0.03$ after approximately 1,000 s (ref. 17). In contrast, the X-ray light curve, begun at 396 s after the GRB owing to observing constraints, shows a steady decay consistent with a single power law with slope $\alpha = 1.36 \pm 0.01$ and no evidence of flares or breaks up to $t > 7 \times 10^5$ s post-burst¹⁴. The absence of any additional emission components from late-time central engine activity superimposed on the afterglow light curve allows a straightforward interpretation of the light curves in the context of current GRB models. The steep-shallow decay of optical emission from GRB 090102 is characteristic of an afterglow whose early-time light is dominated by fading radiation generated in the reverse shock^{9,18}.

Figure 1 shows the RINGO exposure obtained on the night of 2 January 2009. The afterglow of GRB 090102 is clearly visible, as are six

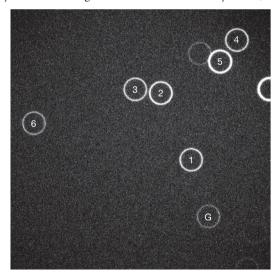


Figure 1 | **RINGO observation of the field of GRB 090102 observed 2009 Jan 2.** The field of view is 4.6×4.6 arcmin. The data have been dark-subtracted and flat-fielded using standard astronomical algorithms. The afterglow of GRB 090102 is labelled G and six foreground sources are labelled 1–6. Foreground source 5 is contaminated by an overlapping faint source, and so was not used in further analysis. We followed our standard RINGO reduction procedure in which flux traces for all objects on all nights were extracted within annuli with inner (8 arcsec) and outer (14 arcsec) radii sufficient to ensure that seeing variations do not influence the extracted fluxes. The traces were then sky-subtracted by the normalized flux inside the inner trace radius and divided through by an average of the traces from routinely obtained zero-polarization standards²⁹ to remove the known 2.7% instrumental polarization. The resulting flux traces for a sample of objects and the GRB are presented in Fig. 2.

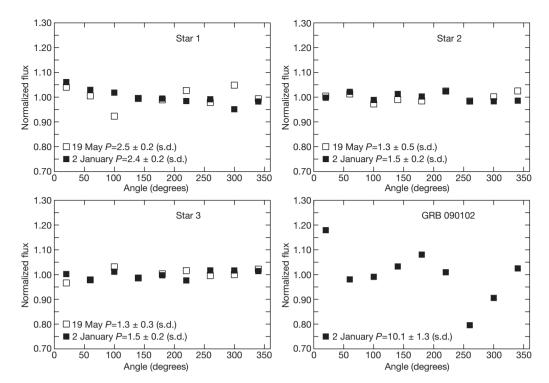


Figure 2 | RINGO data for GRB 090102 and calibration

sources. Shown are example flux traces around the rings of three of the foreground objects (1-3) and GRB 090102, showing a clear $\sin 2\theta$ signal for the GRB. For the foreground objects, traces are presented taken simultaneously with GRB 090102 (filled symbols) and on the night of 19 May 2009 (unfilled symbols). Analysis³⁰ of different subsets 13 of the data in the GRB trace allows a mean polarization and standard deviation to be measured, giving a value of $10.1 \pm 1.3\%$ for GRB 090102. Objects 2 and 3 have low polarization ($\leq 1.5\%$) in both exposures and set limits on uncorrected instrumental polarization effects. Object 1 is detected as weakly polarized (2.5%) in both measurements, demonstrating the stability of the instrumental set-up (the instrument reference position angle varies with the telescope altitude-azimuth mount between the two epochs, so the traces are not in phase, yet the amplitude of variation and hence the derived polarization is similar). Objects 4 and 6 (not plotted) show similar stable weak polarization between different epochs of ~3% and \sim 4% respectively.

brighter foreground objects. Detection of these objects allowed us to perform additional checks on the instrumental calibration at the time of the GRB. In addition, by observing the same region of sky at later dates after the GRB had faded (28 January 2009, 18 April 2009 and 19 May 2009) the stability of RINGO was also verified. The measured optical (4,600–7,200 Å) polarization of GRB 090102 is $P=10.2\pm1.3\%$, in contrast to the foreground objects that show $P\approx1$ –4% (Fig. 2). A simple Monte Carlo analysis (Fig. 3) was performed to estimate the significance of the polarization measurements. This showed that the rank of our GRB measurement amongst a distribution of randomly reordered GRB trace data was 9,988/10,000.

In interpreting our measurement, we first consider whether such a polarization could be produced via the production of magnetic instabilities in the shock front (Fig. 4c). A very optimistic estimate of the coherence length can be made by assuming it grows at about the speed of light in the local fluid frame after the field is generated at the shock front—in this situation, polarized radiation would come from a number of independent ordered magnetic field patches. A measured polarization of 10% is at the very uppermost bound for such a model¹⁹ and therefore seems unlikely. As an alternative to the 'patch' model, we have also considered the case where the observer's line of sight is close to the jet edge²⁰ (Fig. 4b). In this case, because the magnetic fields parallel and perpendicular to the shock front could have significantly different averaged strengths²¹ a polarization signal can also be produced. However, applying this model to GRB 090102 we would have expected a steepening of the light curve (a 'jet-break') just after the time of our polarization measurement, rather than the observed flattening. Similarly, there is no evidence of a jet break in the X-ray light curve up to late times. Furthermore, our detection of 10% is much higher than the reported polarization signal of a few per cent associated with a jet break in the late-time afterglow of other events^{22,23}. We also rule out an Inverse Compton origin for the optical polarization—a mechanism suggested to explain earlier gamma-ray polarization measurements²⁴—in which lower-energy photons are scattered to higher energies by colliding with electrons in the relativistic flow. If Inverse Compton emission is present, it is more likely to contribute to the high-energy X-ray and gamma-ray bands than the optical band and again requires the observer's line of sight to be close to the edge of the jet (Fig. 4b) to produce significant polarization, which, as we have already discussed, is not the case for GRB 090102.

It therefore seems apparent that in the case of GRB 090102 the high polarization signal requires the presence of large-scale ordered magnetic fields in the relativistic outflow (Fig. 4a). As the measurement was obtained while the reverse-shock emission was dominant

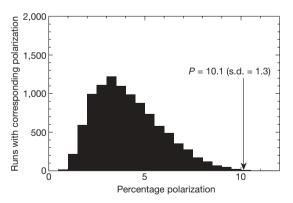
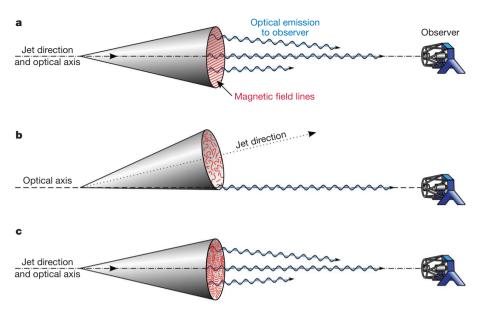


Figure 3 | Monte Carlo simulation using GRB090102 data. The distribution of measured polarizations derived from randomly reordered GRB trace data demonstrates the significance of the actual result. The Monte Carlo N is 10,000, and the mean of the reordered data is 4.2% with an s.d. of 1.7%. The measured value for the GRB (10.1%) is shown with an arrow, and is highly significant (rank = 9,988/10,000). Similar analyses for the foreground objects in the frames confirms that objects 2 (rank 780/1,000) and 3 (rank 540/1,000) have no detectable polarization at the level of 1.5% and that objects 1 (2.5%; rank 969/1,000), 4 (3.3%; rank 927/1,000) and 6 (4.1%; rank 913/1,000) have measured polarizations in line with the expected values for stars within our Galaxy²⁹.



in GRB 090102, the detection of significant polarization provides the first direct evidence that such magnetic fields are present when significant reverse shock emission is produced. Magnetization of the outflow can be expressed as the ratio of magnetic to kinetic energy flux σ . The degree of magnetization cannot be sufficient for the jet to be completely Poynting-flux-dominated (σ >1) because then we would expect it to suppress a reverse shock²⁵. We can therefore reconcile the detection of polarization in GRB 090102 and our previous non-detection in GRB 060418 in a unified manner if GRB jets have magnetization of σ ≈ 1. In the GRB 060418 case, the jet would have had slightly higher magnetization than unity, resulting in the suppression of a reverse shock, while GRB 090102 would have σ slightly smaller than unity, which is optimal to produce bright reverse-shock emission. Of course, owing to the small sample (only two objects), we cannot rule out the possibility that each GRB jet had very different magnetization.

Finally, we note that a high degree of polarization is also predicted for the prompt gamma-ray emission in the presence of large-scale ordered magnetic fields^{26,27}. Recent claims of rapidly ($\sim 10 \, \mathrm{s}$) variable gamma-ray polarization from less than 4% up to 43% ($\pm 25\%$) in the prompt emission of GRB 041219A²⁸ lend further support to models with magnetized outflows and offer the possibility that the peak optical polarization from GRB 0901012 could have been even higher than that measured in our 60-s exposure.

Received 7 September; accepted 16 October.

- Lyutikov, M. The electromagnetic model of gamma ray bursts. N. J. Phys. 8, 119–143 (2006).
- Zhang, B. & Mészáros, P. Gamma-ray bursts: progress, problems and prospects. Int. J. Mod. Phys. A 19, 2385–2471 (2004).
- 3. Piran, T. Gamma ray bursts and the fireball model. *Phys. Rep.* **314**, 575–667 (1999).
- Lyutikov, M. Magnetocentrifugal launching of jets from discs around Kerr black holes. Mon. Not. R. Astron. Soc. 396, 1545–1552 (2009).
- Coburn, W. & Boggs, S. E. Polarization of the prompt γ-ray emission from the γ-ray burst of 6 December 2002. Nature 423, 415–417 (2003).
- 6. Rutledge, R. E. & Fox, D. B. Re-analysis of polarization in the γ -ray flux of GRB 021206. Mon. Not. R. Astron. Soc. **350**, 1288–1300 (2004).
- Willis, D. R. et al. Evidence of polarisation in the prompt gamma-ray emission from GRB 930131 and GRB 960924. Astron. Astrophys. 439, 245–253 (2005).
- McGlynn, S. et al. Polarisation studies of the prompt gamma-ray emission from GRB 041219a using the spectrometer aboard INTEGRAL. Astron. Astrophys. 466, 895–904 (2007).
- Zhang, B., Kobayashi, S. & Mészáros, P. Gamma-ray burst early optical afterglows: implications for the initial Lorentz factor and the central engine. Astrophys. J. 595, 950–954 (2003).
- Zhang, B. & Kobayashi, S. Gamma-ray burst early afterglows: reverse shock emission from arbitrarily magnetized ejecta. Astrophys. J. 628, 315–334 (2005).
- Steele, I. A. et al. RINGO: a novel ring polarimeter for rapid GRB followup. Proc. SPIE 6269, 179S (2006).
- 12. Steele, I. A. et al. performance and first results. Proc. SPIE 5489, 679-692 (2004).

Figure 4 | Competing models of GRB magnetic field structure. The schematic shows three representations of a GRB outflow in the context of the standard fireball model for a variety of magnetic field structures and different orientations to the observer's line of sight (optical axis). A large degree of polarization is predicted when the ejected material is threaded with a largescale ordered magnetic field as shown in a and is the favoured model to explain the measured polarization in GRB 090102. Alternatively, if no ordered magnetic field is present and instead a tangled magnetic field is produced in the shock front, the detected light will be polarized only if the observer's line of sight is close to the jet edge (b). In this case, however, the predicted steepening of the light curve that is expected when observing an off-axis jet is inconsistent with the flattening exhibited in the light curve of GRB 090102. A compromise is shown in c in which the shock front contains a number of independent patches of locally ordered magnetic fields; a measured polarization of 10% is at the very uppermost bound for such a model.

- 13. Mundell, C. G. et al. Early optical polarization of a gamma ray burst afterglow. *Science* 315, 1822–1824 (2007).
- 14. Mangano, V. et al. Swift observations of GRB 090102. GCN Rep. 192.1 (2009).
- Klotz, A. et al. GRB 090102: TAROT Calern Observatory optical observations. GCN Circ. 8761 (2009).
- Covino, S. et al. GRB 090102: REM observations of a bright afterglow. GCN Circ. 8763 (2009).
- 17. Gendre, B. et al. Testing the GRB models with the strange afterglow of GRB090102. Mon. Not. R. Astron. Soc. (submitted).
- Gomboc, A. et al. Multiwavelength analysis of the intriguing GRB 061126: the reverse shock scenario and magnetization. Astrophys. J. 660, 489–495 (2008).
- Gruzinov, A. & Waxman, E. Gamma-ray burst afterglow: polarization and analytical light curves. Astrophys. J. 511, 852–861 (1999).
- 20. Gruzinov, A. Strongly polarized optical afterglows of gamma-ray bursts. *Astrophys. J.* **525**, L29–L31 (1999).
- Medvedev, M. V. & Loeb, A. Generation of magnetic fields in the relativistic shock of gamma-ray burst sources. Astrophys. J. 526, 697–706 (1999).
- Covino, S. et al. GRB 990510: linearly polarized radiation from a fireball. Astron. Astrophys. 348, L1–L4 (1999).
- 23. Wijers, R. A. M. et al. Detection of polarization in the afterglow of GRB 990510 with the ESO Very Large Telescope. *Astrophys. J.* **523**, L33–L36 (1999).
- Lazzati, D., Rossi, E., Ghisellini, G. & Rees, M. J. Compton drag as a mechanism for very high linear polarization in gamma-ray bursts. *Mon. Not. R. Astron. Soc.* 347, L1–L5 (2004).
- 25. Mimica, P., Giannios, D. & Aloy, M. A. Deceleration of arbitrarily magnetized GRB ejecta: the complete evolution. *Astron. Astrophys.* **494**, 879–890 (2009).
- Fan, Y. Interpretation and implications of the non-detection of GeV spectrum excess by the Fermi Gamma-ray Space Telescope in most gamma-ray bursts. Mon. Not. R. Astron. Soc. 397, 1539–1548 (2009).
- 27. Granot, J. The most probable cause for the high gamma-ray polarization in GRB 021206. *Astrophys. J.* **596**, L17–L21 (2003).
- Götz, D., Laurent, P., Lebrun, F., Daigne, F. & Bošnjak, Ž. Variable polarization measured in the prompt emission of GRB 041219A using IBIS on board INTEGRAL. Astrophys. J. 695, L208–L212 (2009).
- Schmidt, G. D., Elston, R. & Lupie, O. L. The Hubble Space Telescope northernhemisphere grid of stellar polarimetric standards. Astron. J. 104, 1563–1567 (1992).
- Clarke, D. & Neumayer, D. Experiments with a novel CCD stellar polarimeter. Astron. Astrophys. 383, 360–366 (2002).

Acknowledgements The Liverpool Telescope is operated on the island of La Palma by Liverpool John Moores University in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofisica de Canarias with financial support from the UK Science and Technology Facilities Council. C.G.M. acknowledges financial support from the Royal Society and Research Councils UK. We thank J. Marchant for preparing Fig. 4 for us.

Author Contributions I.A.S. designed and built the instrument, and did data reduction and analysis. C.G.M. developed the initial scientific justification for the instrument and interpreted results. R.J.S. did data reduction and analysis. S.K. did theoretical interpretation of results. C.G. coordinated observations and identified the afterglow.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.A.S. (ias@astro.livjm.ac.uk).

A connection between star formation activity and cosmic rays in the starburst galaxy M82

The VERITAS Collaboration*

Although Galactic cosmic rays (protons and nuclei) are widely believed to be mainly accelerated by the winds and supernovae of massive stars, definitive evidence of this origin remains elusive nearly a century after their discovery¹. The active regions of starburst galaxies have exceptionally high rates of star formation, and their large size—more than 50 times the diameter of similar Galactic regions—uniquely enables reliable calorimetric measurements of their potentially high cosmic-ray density². The cosmic rays produced in the formation, life and death of massive stars in these regions are expected to produce diffuse γ-ray emission through interactions with interstellar gas and radiation. M82, the prototype small starburst galaxy, is predicted^{3,4} to be the brightest starburst galaxy in terms of γ -ray emission. Here we report the detection of >700-GeV γ-rays from M82. From these data we determine a cosmic-ray density of 250 eV cm⁻³ in the starburst core, which is about 500 times the average Galactic density. This links cosmic-ray acceleration to star formation activity, and suggests that supernovae and massive-star winds are the dominant accelerators.

M82 is a bright galaxy located approximately 12,000,000 light yr from Earth, in the direction of the Ursa Major constellation⁵. For hundreds of millions of years, M82 has been gravitationally interacting with nearby galaxies, including the larger spiral galaxy M81 (ref. 6). Over time, interactions with these neighbours have deformed M82, creating an active starburst region in its centre with a diameter of ~1,000 light yr (ref. 7). The NASA Hubble Space Telescope reveals hundreds of young, massive (10⁴–10⁶ solar masses) clusters in this starburst region8. Throughout this compact region, stars are being formed at a rate approximately ten times faster than in entire 'normal' galaxies like the Milky Way, and the supernovae rate is 0.1–0.3 yr⁻¹ (refs 9, 10). The intense radio synchrotron emission observed in the central region of M82 suggests a very high cosmic-ray energy density, about two orders of magnitude higher than in the Milky Way¹¹. The region also contains a high mean (molecular) gas density, of about 150 particles per cubic centimetre, or about 10⁹ solar masses in total¹². Given the high cosmic-ray and gas densities, M82 has long been viewed as a promising target for γ -ray observatories⁷. However, emission from it was not detected above 100 MeV by the NASA Energetic Gamma-Ray Experiment Telescope experiment¹³, nor during previous very-high-energy (VHE, energy >100 GeV) γ -ray observations of M82 made using the Whipple 10-m Telescope¹⁴ and by the High Energy Gamma Ray Astronomy¹⁵ experiment. The latter two set upper limits on the flux from M82 at \sim 10% of that from the Crab Nebula, the brightest steady VHE source in the sky. These limits are well above the sensitivity of the Very Energetic Radiation Imaging Telescope Array System (VERITAS).

VERITAS¹⁶ is located in southern Arizona and has been fully operational since September 2007. It consists of a stereoscopic array of four 12-m-diameter optical telescopes equipped with sensitive cameras (3.5°) field of view) that detect short $(\sim 3-ns)$ flashes of

ultraviolet and blue light known as Cherenkov radiation. This light is emitted in the electromagnetic cascade of secondary particles resulting from the interaction of a VHE γ -ray in the upper atmosphere. VERITAS has an energy threshold of ~ 100 GeV, an energy resolution of $\sim 15\%$ and an angular resolution of $\sim 0.1^{\circ}$ per event.

We observed M82 using VERITAS for a total of \sim 137 h of quality-selected live time between January 2008 and April 2009 at a mean zenith angle of 39°. This exceptionally long exposure was made entirely during periods of astronomical darkness and clear atmospheric conditions. The analysis of these data was performed according to the standard VERITAS analysis procedure ¹⁷ using event-selection criteria optimized a priori for low-flux, hard-spectrum sources. We observed an excess of 91 γ -ray-like events (\sim 0.7 photons per hour) above the estimated background (267 events) from the direction of M82 (Supplementary Information). This excess corresponds to a post-trial statistical significance of 4.8 σ , or a chance probability of 7.7 \times 10⁻⁷, and represents the discovery of VHE γ -ray emission from M82 (Fig. 1). The observed

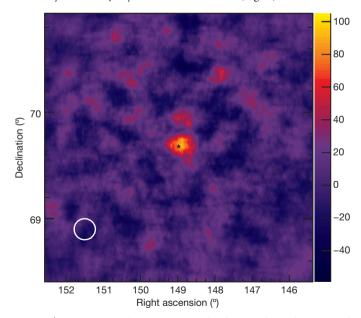


Figure 1 | VHE image of the M82 region. The sky map shows the measured excess (colour scale) of γ -ray-like events above the estimated background from a region centred on M82. Each pixel contains the excess in a circular region of radius 0.1°. The map is oversampled; neighbouring pixels are thus correlated. The background for each point is estimated using an annulus centred on its position (the ring method²8). The spatial distribution of the observed excess is consistent with that expected from a point-like source located near the core of M82. The white circle represents the VERITAS point spread function (68% containment) for individual γ -rays. The uncertainty in the source localization is much smaller. The black star denotes the location of the core of M82. The coordinates are for the J2000 epoch.

differential VHE γ -ray spectrum (Fig. 2) is best fitted using a power-law function with a photon index of $\Gamma=2.5\pm0.6_{\text{stat}}\pm0.2_{\text{syst}}$ (the two uncertainties corresponding to statistical and systematic errors, respectively). The measured γ -ray flux is $(3.7\pm0.8_{\text{stat}}\pm0.7_{\text{syst}})\times10^{-13}\,\text{cm}^{-2}\,\text{s}^{-1}$ above the 700-GeV energy threshold of the analysis, and no flux variations are observed. The luminosity of M82 above 700 GeV that we infer from the γ -ray flux is $2\times10^{32}\,\text{W}$, which is about 2×10^6 times smaller than its far-infrared (100- μ m) luminosity 18 .

Having a flux of 0.9% of that observed from the Crab Nebula, M82 is among the weakest VHE sources ever detected. Although VERITAS has detected several confirmed VHE sources with flux values close to this, we performed a large number of tests to ensure that systematic effects could not potentially create a spurious signal in the data (Supplementary Information). None of these tests gave any indication that the observed signal is an artefact.

Before our discovery of VHE γ -ray emission from M82 using VERITAS, all known extragalactic VHE sources were clearly associated with an active galactic nucleus, an object powered by accretion onto a supermassive black hole. Although M82 may have a supermassive black hole at its centre, it exhibits at most only a weak level of activity that would point to its being an active galactic nucleus¹⁹. However, the high rate of star formation in M82 implies the presence of numerous strong shock waves in supernova remnants and around massive young stars. In the Milky Way, similar shock waves are known to accelerate electrons to very high energies, and they are suspected to similarly accelerate ions. This acceleration is expected to supply the cosmic rays that permeate both the Galaxy and M82, and which produce diffuse γ -ray emission.

The most recent theoretical models^{2–4,7} predict a VHE γ -ray flux from M82 on the basis of the acceleration and propagation of cosmic rays in the starburst core. The various calculated fluxes are all close to the value measured by VERITAS. Using the model³ shown in Fig. 2,

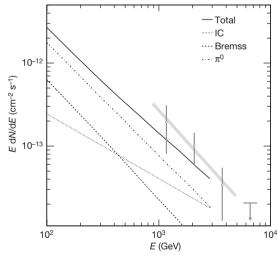


Figure 2 | Gamma-ray flux compared with a theoretical prediction. The differential energy spectrum (EdN/dE, where E denotes energy and N is the number of photons) of M82 observed using VERITAS between ~0.9 TeV and \sim 5 TeV. The data are given by open diamonds with 1σ statistical error bars, and can be fitted ($\chi^2 = 0.1$ with 1 d.f.) with a power-law function (thick grey line: $dN/dE \approx E^{-\Gamma}$, where E is measured in teraelectronvolts and Γ = 2.5 \pm 0.6_{stat} \pm 0.2_{syst}). The VERITAS flux upper limit (99% confidence level²⁹) shown at \sim 6.6 TeV is above the extrapolation of the fitted power-law function at these energies. The thin lines represent a recent model3 for the γ -ray emission from M82. The thin solid line is the total emission predicted and the dashed lines represent components of this emission that result from the interactions of cosmic-ray ions with interstellar matter (decay of neutral pions (π^0)), from radiation from cosmic-ray electrons through inverse Compton scattering (IC) and from Bremsstrahlung radiation (Bremss). The IC and π^0 decay components are the dominant contributions of cosmic-ray electrons and ions, respectively. Notably, the spectral slopes of these dominant components are markedly different.

from the VHE flux we estimate the cosmic-ray density in the starburst core of M82 to be \sim 250 eV cm⁻³, which is approximately 500 times the average Milky Way density. Although the cosmic-ray density of the M82 core is significantly higher, the total cosmic-ray energy content of the two systems is similar because in terms of volume the Milky Way is about 500 times the larger. The lifetime of cosmic-ray particles in the M82 core is constrained to approximately 1,000,000 yr owing to energy losses through adiabatic cooling in the starburst wind and through collisions with interstellar gas nuclei. This is about 30 times shorter than the lifetime of the gigaelectronvolt-band particles in the Milky Way, which dominate the local cosmic-ray density. Thus, a correspondingly larger source power is needed to replenish these particles in M82 to maintain similar cosmic-ray energy content. Interestingly, the estimated supernova rate in M82 is about a factor of 30 larger than in the Milky Way. The VERITAS data therefore show an enhancement in the cosmic-ray acceleration that matches the enhancement in energy input by massive stars and supernovae. This correlation strongly supports the long-held theory that these objects have a dominant role in cosmic-ray production.

Although the VERITAS data strongly indicate that smaller shocks (for example those in supernova remnants) are the predominant cosmic-ray acceleration sites, it cannot be ruled out that this acceleration occurs on larger (>30-light-yr) scales in a more distributed fashion¹. Significantly lower estimates of the M82 supernova rate⁴ would also suggest other potential sources of cosmic-ray acceleration. However, alternative sources of mechanical energy for cosmic-ray acceleration, such as galactic rotation¹, can be ruled out.

The aforementioned theoretical models include significant contributions from both leptonic (for example electron) and hadronic (for example ion) particle interactions, which are expected to give different VHE γ -ray spectra (Fig. 2). Cosmic-ray ions create VHE γ -rays through collisions with interstellar matter. This process creates unstable particles called pi mesons (pions). Electrically neutral pions directly decay into γ-rays. Charged pions eventually decay into neutrinos and electrons. The latter emit synchrotron radiation in the radio and infrared bands through interactions with the ambient magnetic field. The radio emission from these secondary electrons can be used to place an upper limit on the γ -ray flux produced by cosmic-ray ions, thus helping to further discriminate between VHE γ -rays emitted by cosmic-ray ions and those coming from cosmic-ray electrons. The radio flux observed at a frequency of 32 GHz (ref. 20) implies that cosmic-ray ions would not produce a γ-ray flux at 20 GeV greater than about 2.5×10^{-9} cm⁻² s⁻¹, unless the magnetic field in M82 is considerably weaker than the conventional estimate, of 8 nT. An extrapolation of the VHE γ -ray spectrum measured with VERITAS using the fitted power-law index, $\Gamma = 2.5$, would exceed that limit by a factor of two, whereas an extrapolation with $\Gamma = 2.3$, which is within the uncertainty range of the fitted value, would satisfy the limit. The comparison suggests that either the true γ -ray spectrum between 10 GeV and 1 TeV is slightly harder (has a lower photon index) than our best-fit spectrum suggests, or the γ-ray emission does not come predominantly from cosmic-ray ions.

The observed radio emission may also come from the relativistic cosmic-ray electrons accelerated in M82. All electrons interact with ambient infrared photons, boosting them into the hard X-ray/soft γ -ray band by means of inverse Compton scattering. This non-thermal process contributes $\sim 25\%$ to the diffuse X-ray flux²¹, the remainder of which originates from thermal emission of hot gas. Observational limits on the steady, non-thermal diffuse X-ray emission place its luminosity, at a photon energy of 5 keV, not significantly higher than the VHE γ -ray luminosity observed using VERITAS. These X-ray data provide a lower limit on the amplitude of the interstellar magnetic field that is about a third of the current estimate (8 nT) and, hence, an upper limit on the absolute number of relativistic cosmic-ray electrons in M82 with kinetic energies of ~ 1 GeV. Electrons of much higher kinetic energy (~ 10 TeV) are needed to produce VHE γ -rays through inverse Compton scattering of ambient infrared photons. Both theoretical considerations and a

comparison of the observed VHE γ -ray flux with limits on the cosmic-ray-induced X-ray flux suggest that the inverse Compton emission should have a hard spectrum with a power-law index of $\sim\!\!2$ between 100 keV and 100 GeV. Because $>\!\!100\text{-GeV}$ electrons quickly lose their energy by inverse Compton scattering and synchrotron emission, eventually preventing their further acceleration above a characteristic energy, the inverse Compton radiation spectrum should also steepen and eventually show a cut-off. The identification of a cut-off in this spectrum, potentially observable by combining data from VERITAS and the NASA Fermi Gamma-ray Space Telescope, could demonstrate which type of cosmic-ray particle is responsible for the VHE emission.

The VERITAS measurements of M82 also have implications for the interpretation of the striking correlation observed between the far-infrared emission (from warm dust) and the radio emission (from synchrotron radiation of cosmic-ray electrons) in starburst galaxies 22,23 . Massive-star formation is generally accepted as the origin of both 24 , but consensus is lacking on how such a tight correlation is produced $^{25-27}$. The VHE flux measured from M82 places these models on a sound footing by providing an independent estimate of the cosmic-ray density. The observed VHE flux also requires a hard cosmic-ray spectrum (scaling as the *p*th power of energy, with *p* between -2.1 and -2.3 (refs 3, 4)), which places new constraints on models of the radio/far infrared correlation.

Received 6 July; accepted 30 September 2009. Published online 1 November 2009.

- Butt, Y. Beyond the myth of the supernova-remnant origin of cosmic rays. Nature 460, 701–704 (2009).
- Pohl, M. On the predictive power of the minimum energy condition. 2: Fractional calorimeter behaviour in the diffuse high energy gamma emission of spiral galaxies. Astron. Astrophys. 287, 453–462 (1994).
- Persic, M., Rephaeli, Y. & Arieli, Y. Very-high-energy emission from M 82. Astron. Astrophys. 486, 143–149 (2008).
- de Cea del Pozo, E. Torres, D. F. & Marrero, A. Y. R. Multimessenger model for the starburst galaxy M82. Astrophys. J. 698, 1054–1060 (2009).
- Sakai, S. & Madore, B. F. Detection of the red giant branch stars in M82 using the Hubble Space Telescope. Astrophys. J. 526, 599–606 (1999).
- 6. Yun, M. S., Ho, P. T. P. & Lo, K. Y. A high-resolution image of atomic hydrogen in the M81 group of galaxies. *Nature* **372**, 530–532 (1994).
- Völk, H. J., Aharonian, F. A. & Breitschwerdt, D. The nonthermal energy content and gamma-ray emission of starburst galaxies and clusters of galaxies. Space Sci. Rev. 75, 279–297 (1996).
- 8. Melo, V. P., Muñoz-Tuñón, C., Maíz-Apellániz, J. & Tenorio-Tagle, G. Young super star clusters in the starburst of M82: the catalog. *Astrophys. J.* **619**, 270–290 (2005).
- Kronberg, P. P., Biermann, P. & Schwab, F. R. The nucleus of M82 at radio and X-ray bands – discovery of a new radio population of supernova candidates. Astrophys. J. 291, 693–707 (1985).
- Fenech, D. M., Muxlow, T. W. B., Beswick, R. J., Pedlar, A. & Argo, M. K. Deep MERLIN 5 GHz radio imaging of supernova remnants in the M82 starburst. *Mon. Not. R. Astron. Soc.* 391, 1384–1402 (2008).
- Rieke, G. H., Lebofsky, M. J., Thompson, R. I., Low, F. J. & Tokunaga, A. T. The nature of the nuclear sources in M82 and NGC 253. Astrophys. J. 238, 24–40 (1980).
- Weiß, A., Neininger, N., Hüttemeister, S. & Klein, U. The effect of violent star formation on the state of the molecular gas in M 82. Astron. Astrophys. 365, 571–587 (2001).
- Blom, J. J., Paglione, T. A. D. & Carramiñana, A. Diffuse gamma-ray emission from starburst galaxies and M31. Astrophys. J. 516, 744–749 (1999).
- Nagai, T. Search for Tev Gamma-Ray Emission from Nearby Starburst Galaxies. PhD thesis, Univ. Utah (2005).
- Götting, N. Nachweis von TeV-Gamma-Strahlung aus der Richtung der Blazare H1426+428 und 1ES1959+650 sowie der Radiogalaxie M87 mit den HEGRA-Cherenkov-Teleskopen. PhD thesis, Univ. Hamburg (2007).
- Holder, J. et al. Status of the VERITAS Observatory. AIP Conf. Proc. 1085, 657–660 (2008).
- Daniel, M. K. in Proc. 30th Internat. Cosmic Ray Conf. Vol. 3 (eds Caballero, R. et al.) 1325–1328 (2008).
- Sanders, D. B., Mazzarella, J. M., Kim, D.-C., Surace, J. A. & Soifer, B. T. The IRAS revised bright galaxy sample. Astron. J. 126, 1607–1664 (2003).
- Wills, K. A., Pedlar, A., Muxlow, T. W. B. & Stevens, I. R. A possible active galactic nucleus in M82? N. Astron. Rev. 43, 633–637 (1999).
- Klein, U., Wielebinski, R. & Morsi, H. W. Radio continuum observations of M82. Astron. Astrophys. 190, 41–46 (1988).
- 21. Strickland, D. K. & Heckman, T. M. Iron line and diffuse hard X-Ray emission from the starburst galaxy M82. *Astrophys. J.* **658**, 258–281 (2007).
- van der Kruit, P. C. Observations of core sources in Seyfert and normal galaxies with the Westerbork synthesis radio telescope at 1415 MHz. Astron. Astrophys. 15, 110–122 (1971).

- 23. Helou, G., Soifer, B. T. & Rowan-Robinson, M. Thermal infrared and nonthermal radio remarkable correlation in disks of galaxies. *Astrophys. J.* 298, L7–L11 (1985).
- Condon, J. J. Radio emission from normal galaxies. Annu. Rev. Astron. Astrophys. 30, 575–611 (1992).
- 25. Voelk, H. J. The correlation between radio and far-infrared emission for disk galaxies a calorimeter theory. *Astron. Astrophys.* **218**, 67–70 (1989).
- 26. Helou, G. & Bicay, M. D. A physical model of the infrared-to-radio correlation in galaxies. *Astrophys. J.* 415, 93–100 (1993).
- 27. Groves, B. A., Cho, J., Dopita, M. & Lazarian, A. The radio-FIR correlation: is MHD turbulence the cause? *Publ. Astron. Soc. Aust.* 20, 252–256 (2003).
- 28. Berge, D., Funk, S. & Hinton, J. Background modelling in very-high-energy γ -ray astronomy. *Astron. Astrophys.* **466**, 1219–1229 (2007).
- 29. Helene, O. Upper limit of peak area. *Nucl. Instrum. Methods Phys. Res.* 212, 319–322 (1983)

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research is supported by grants from the US Department of Energy, the US National Science Foundation and the Smithsonian Institution, and by the National Science and Engineering Research Council of Canada, Science Foundation Ireland and the UK Science and Technology Facilities Council. We acknowledge the excellent work of the technical support staff at the Fred Lawrence Whipple Observatory and the institutions that collaborated in the construction and operation of the VERITAS array.

Author Contributions VERITAS is a collaboration of scientists who jointly participate in all aspects of the scientific effort: the collecting of data, the development of software for analysis and simulations, the analysis of data and the interpretation of the results. Every author has read the paper and agrees with the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.B. (wbenbow@cfa.harvard.edu).

The VERITAS Collaboration

V. A. Acciari^{1,23}, E. Aliu², T. Arlen³, T. Aune⁴, M. Bautista⁵, M. Beilicke⁶, W. Benbow¹, D. Boltuch², S. M. Bradbury⁷, J. H. Buckley⁶, V. Bugaev⁶, K. Byrum⁸, A. Cannon⁹, O. Celik³, A. Cesarini¹⁰, Y. C. Chow³, L. Ciupik¹¹, P. Cogan⁵, P. Colin¹², W. Cui¹³, R. Dickherber⁶, C. Duke¹⁴, S. J. Fegan³, J. P. Finley¹³, G. Finnegan¹², P. Fortin¹⁵, L. Fortson¹¹, A. Furniss⁴, N. Galante¹, D. Galli¹³, K. Gibbs¹, G. H. Gillanders¹⁰, S. Godambe¹², J. Grube⁹, R. Guenette⁵, G. Gyuk¹¹, D. Hanna⁵, J. Holder², D. Horan¹⁶, C. M. Hui¹², T. B. Humensky¹⁷, A. Imran¹⁸, P. Kaaret¹⁹, N. Karlsson¹¹, M. Kertzman²⁰, D. Kieda¹², J. Kildea¹, A. Konopelko²¹, H. Krawczynski⁶, F. Krennrich¹⁸, M. J. Lang¹⁰, S. LeBohec¹², G. Maier⁵, S. McArthur⁶, A. McCann⁵, M. McCutcheon⁵, J. Millis²², P. Moriarty²³, R. Mukherjee¹⁵, T. Nagai¹⁸, R. A. Ong³, A. N. Otte⁴, D. Pandel¹⁹, J. S. Perkins¹, F. Pizlo¹³, M. Pohli⁸, J. Quinn⁹, K. Ragan⁵, L. C. Reyes²⁴, P. T. Reynolds²⁵, E. Roache¹, H. J. Rose⁷, M. Schroedter¹⁸, G. H. Sembroski¹³, A. W. Smith⁸, D. Steele¹¹, S. P. Swordy¹⁷, M. Theiling¹, S. Thibadeau⁶, A. Varlotta¹³, V. V. Vassiliev³, S. Vincent¹², R. G. Wagner⁸, S. P. Wakely¹⁷, J. E. Ward⁹, T. C. Weekes¹, A. Weinstein³, T. Weisgarber¹⁷, D. A. Williams⁴, S. Wissel¹⁷, M. Wood³ & B. Zitzer¹³

¹Fred Lawrence Whipple Observatory, Harvard-Smithsonian Center for Astrophysics, Amado, Arizona 85645, USA. ²Department of Physics and Astronomy and the Bartol Research Institute, University of Delaware, Newark, Delaware 19716, USA. ³Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA. ⁴Santa Cruz Institute for Particle Physics and Department of Physics, University of California, Santa Cruz, California 95064, USA. 5Physics Department, McGill University, Montreal, Quebec H3A 2T8, Canada. ⁶Department of Physics, Washington University, St Louis, Missouri 63130, USA. ⁷School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK. 8 Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. ⁹School of Physics, University College Dublin, Belfield, Dublin 4, Ireland. 10 School of Physics, National University of Ireland, Galway, Ireland. 11 Astronomy Department, Adler Planetarium and Astronomy Museum, Chicago, Illinois 60605, USA. ¹²Department of Physics and Astronomy, University of Utah, Salt Lake City, Utah 84112, USA. ¹³Department of Physics, Purdue University, West Lafayette, Indiana 47907, USA. ¹⁴Department of Physics, Grinnell College, Grinnell, Iowa 50112-1690, USA. ¹⁵Department of Physics and Astronomy, Barnard College, Columbia University, New York, New York 10027, USA. ¹⁶Laboratoire Leprince-Ringuet, Ecole Polytechnique, CNRS/IN2P3, F-91128 Palaiseau, France. ¹⁷Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA. ¹⁸Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA. ¹⁹Department of Physics and Astronomy, University of Iowa, Van Allen Hall, Iowa City, Iowa 52242, USA. 20 Department of Physics and Astronomy, DePauw University, Greencastle, Indiana 46135-0037, USA. ²¹Department of Physics, Pittsburg State University, 1701 South Broadway, Pittsburg, Kansas 66762, USA. ²²Department of Physics, Anderson University, 1100 East 5th Street, Anderson, Indiana 46012, USA. ²³Department of Life and Physical Sciences, Galway-Mayo Institute of Technology, Dublin Road, Galway, Ireland. ²⁴Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA. ²⁵Department of Applied Physics and Instrumentation, Cork Institute of Technology, Bishopstown, Cork, Ireland.

Disordered, quasicrystalline and crystalline phases of densely packed tetrahedra

Amir Haji-Akbari¹*, Michael Engel¹*, Aaron S. Keys¹, Xiaoyu Zheng³, Rolfe G. Petschek⁵, Peter Palffy-Muhoray⁴ & Sharon C. Glotzer^{1,2}

All hard, convex shapes are conjectured by Ulam to pack more densely than spheres1, which have a maximum packing fraction of $\phi = \pi/\sqrt{18} \approx 0.7405$. Simple lattice packings of many shapes easily surpass this packing fraction^{2,3}. For regular tetrahedra, this conjecture was shown to be true only very recently; an ordered arrangement was obtained via geometric construction with $\phi = 0.7786$ (ref. 4), which was subsequently compressed numerically to ϕ = 0.7820 (ref. 5), while compressing with different initial conditions led to $\phi = 0.8230$ (ref. 6). Here we show that tetrahedra pack even more densely, and in a completely unexpected way. Following a conceptually different approach, using thermodynamic computer simulations that allow the system to evolve naturally towards highdensity states, we observe that a fluid of hard tetrahedra undergoes a first-order phase transition to a dodecagonal quasicrystal⁷⁻¹⁰, which can be compressed to a packing fraction of $\phi = 0.8324$. By compressing a crystalline approximant of the quasicrystal, the highest packing fraction we obtain is $\phi = 0.8503$. If quasicrystal formation is suppressed, the system remains disordered, jams and compresses to $\phi = 0.7858$. Jamming and crystallization are both preceded by an entropy-driven transition from a simple fluid of independent tetrahedra to a complex fluid characterized by tetrahedra arranged in densely packed local motifs of pentagonal dipyramids that form a percolating network at the transition. The quasicrystal that we report represents the first example of a quasicrystal formed from hard or non-spherical particles. Our results demonstrate that particle shape and entropy can produce highly complex, ordered

The packing of shapes has drawn the attention of humankind since ancient times. A Sanskrit work from 499 AD reveals the first-known mathematical study of the face-centred cubic arrangement of spheres 11. Kepler conjectured and Hales only recently proved the sphere-close-packing fraction of $\phi=\pi/\sqrt{18}\approx 0.7405$ achieved by face-centred cubic and its stacking variations 12. Much less is known about the packing of other shapes. In the case of ellipsoids, periodic arrangements were found with packing fractions up to $\phi=0.7707$ (ref. 3). The recent progress in the synthesis of non-spherical particles of sizes ranging from nanometres to micrometres 13 has focused attention on the problem of packing three-dimensional shapes such as tetrahedra 14.

In hard particle systems, the potential energy of two particles is considered infinite if they interpenetrate and zero otherwise. All permissible configurations of such systems have the same energy, so the equilibrium structure at constant volume maximizes entropy. Surprisingly, hard particles can maximize entropy by ordering. Entropydriven disorder—order transitions (first predicted by Onsager¹⁵ for hard thin rods and Kirkwood¹⁶ for hard spheres) are now well established for the originally controversial case of spheres, as well as for rods¹⁷,

ellipsoids¹⁸ and other shapes^{19,20}. In the limit of infinite pressure, an arrangement with maximum packing fraction is stable because it minimizes specific volume and Gibbs free energy.

One of the simplest shapes for which the packing problem is still unsolved is the regular tetrahedron. Tetrahedra do not tile Euclidean space. However, if extra space is allowed between tetrahedra, or between groups of tetrahedra, dense ordered structures become possible. Imagine building a dense cluster, one tetrahedron at a time. As shown in Fig. 1A, a pentagonal dipyramid is easily built from five tetrahedra if one allows an internal gap of 7.36°. Two pentagonal dipyramids can share a single tetrahedron to form a nonamer. Twelve interpenetrating pentagonal dipyramids define an icosahedron with a gap of 1.54 steradians. In Fig. 1A, tetrahedral dice are stuck together with modelling putty, which distributes the gap that would be present in each motif if most of the adjacent faces were touching. Pentagonal dipyramids and icosahedra are locally dense, but exhibit non-crystallographic symmetries. The problem of extending or arranging them into space-filling bulk structures is non-trivial. For example, adding a second shell to the icosahedron generates a larger cluster with icosahedral symmetry and 70 tetrahedra, but decreases the packing fraction. For later use, we introduce here a dense, one-dimensional packing given by a linear arrangement of tetrahedra with touching faces known as a tetrahelix, or Bernal spiral.

Recent theoretical works have discussed possible ordered phases of hard tetrahedra formed by some of these motifs. Conway and Torquato²¹ proposed the Scottish, Irish and Welsh configurations, derived from the polytetrahedral networks of clathrate hydrates with packing fractions of up to $\phi = 0.7175$. Chen⁴ constructed a crystalline structure formed from nonamers with $\phi = 0.7786$, the first to exceed sphere close packing, and showing that tetrahedra obey Ulam's conjecture. Chen's structure was subsequently compressed to $\phi = 0.7820$ (ref. 5) and the same algorithm applied to different initial conditions yielded $\phi = 0.8230$ (ref. 6). All these packings originated from either geometric considerations or numerical compression. No simulation or experiment has yet reported the spontaneous formation of an ordered phase of hard tetrahedra. Aside from studies of packing, hard tetrahedra have been used to model the structure of water²². Polytetrahedral networks of atoms are characteristic for Frank-Kasper phases²³, common in intermetallic compounds.

To obtain dense packings of hard regular tetrahedra, we carry out Monte Carlo simulations. Figure 1B–D shows the densest configuration ($\phi=0.8324$) we obtained: we equilibrated an initially disordered fluid of 13,824 tetrahedra at constant $\phi=0.5$ and then we compressed the ordered structure that forms. As demonstrated below, this structure is a quasicrystal, with a packing fraction much greater than all previously proposed arrangements of regular tetrahedra. First, we discuss the thermodynamics of the hard tetrahedron fluid.

¹Department of Chemical Engineering, ²Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA. ³Department of Mathematical Sciences, ⁴Liquid Crystal Institute, Kent State University, Kent, Ohio 44242, USA. ⁵Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106, USA. *These authors contributed equally to this work.

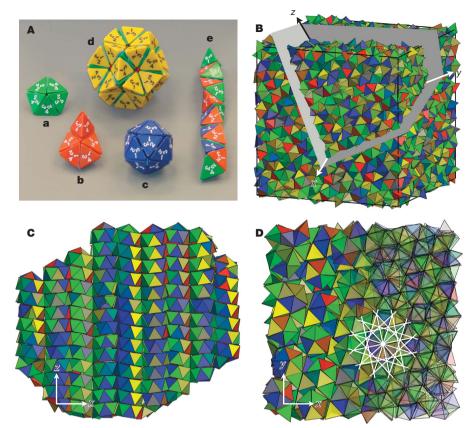


Figure 1 | Packings of tetrahedra obtained by hand and by computer simulation. A, Ideal local packing motifs built from tetrahedral dice stuck together with modelling putty. The pentagonal dipyramid (a), the nonamer (b) and the icosahedron (c) maximize local packing density. The icosahedron can be extended by adding a second shell (d), but then the large gaps between the outer tetrahedra lower the density. The tetrahelix (e) maximizes packing density in one dimension. B–D, A quasicrystal with packing fraction $\phi = 0.8324$ obtained by first equilibrating an initially

disordered fluid of 13,824 hard tetrahedra using Monte Carlo simulation and subsequent numerical compression. The images show an opaque view of the system ($\bf B$) and opaque and translucent views of two rotated narrow slices $\bf C$ and $\bf D$. The white overlay in $\bf D$ shows the distinctive 12-fold symmetry of the dodecagonal quasicrystal. Corrugated layers with normals along the z axis are apparent in $\bf C$. The colouring of the tetrahedra is based on orientation.

Figure 2a shows the equation of state $\phi(P^*)$ obtained from simulations of a small system with 512 tetrahedra and a larger system with 4,096 tetrahedra. Here, $P^* = P\sigma^3/k_BT$ is the reduced pressure and σ the edge length of a tetrahedron. For the small system, the equilibrium packing fraction exhibits an S-shaped transition at $P^* = 58$ and $\phi = 0.47$ from a simple fluid to a more complex fluid, discussed below. At higher pressure the system jams (Supplementary Figs 1 and 2) and, when compressed to nearly infinite pressure, attains a maximum packing fraction of $\phi = 0.7858$. The large system undergoes a first-order transition on compression of the fluid phase and forms a quasicrystal. In Fig. 2b, we analyse the system for the presence of locally dense motifs introduced in Fig. 1A. We see that the fraction of tetrahedra belonging to at least one pentagonal dipyramid increases well before jamming or crystallization. With increasing pressure, interpenetrating pentagonal dipyramids form icosahedra and finally merge into a percolating pentagonal dipyramid network (Fig. 2c, d) as the fraction of tetrahedra in pentagonal dipyramids approaches unity. For the large system, the fraction of tetrahedra in icosahedra suddenly drops at $P^* = 62$, when crystallization occurs. Comparison with the glass shows that far fewer icosahedra remain in the quasicrystal. Figure 2c and d suggests a percolation transition of the pentagonal dipyramid network in both systems at $P_{p}^{*} = 58 \pm 2$, before both jamming and crystallization. We do not observe tetrahedratic liquid crystal phases, which have been suggested by theory²⁴.

Structural changes in the fluid are revealed by the unusual behaviour of its radial distribution function g(r), as shown in Fig. 2e. We find that the first peak near $r = 0.75\sigma$ disappears upon compression

at low pressure, only to reappear for higher pressure, splitting into two peaks at $r = 0.55\sigma$ and $r = 0.80\sigma$. The positions of these peaks are characteristic of face-to-face and edge-to-edge arrangements, respectively, within a single pentagonal dipyramid. This initial loss of structure with increasing pressure or packing fraction is strikingly different from the well-known behaviour of the hard sphere system depicted in Fig. 2f, and underscores the influence of shape in dense packings.

The spontaneous formation of a quasicrystal from the fluid is remarkable since all previously observed crystalline structures of hard particles have unit cells consisting of only a few particles 17,19,20 . From Fig. 1C it can be seen that the quasicrystal consists of a periodic stack of corrugated layers with spacing 0.93σ . The view along the direction of the stacking vector (Fig. 1D) reveals details of the structure within the layers. Twelve-fold symmetric rings formed by interpenetrating tetrahelices exist throughout the structure. The helix chirality is switched by 30° rotations, lowering the symmetry and resulting in a generalized point group of D_{6d} (ref. 25).

The structure of the quasicrystal can be understood more easily by examining the dual representation constructed by connecting the centres of mass of neighbouring tetrahedra. In the dual representation, pentagonal dipyramids are represented by pentagons. The mapping is applied to a layer of an 8,000-particle quasicrystal in Fig. 3a. Recurring motifs are rings of twelve tetrahedra that are stacked periodically to form 'logs' (Fig. 3b), similar to the hexagonal antiprismatic clusters in the tantalum–tellurium system¹⁰. As indicated in Fig. 3a, the symmetry axes of the logs arrange into a non-repeating pattern of squares and triangles (tile edge length 1.83σ)—an observation that we confirm

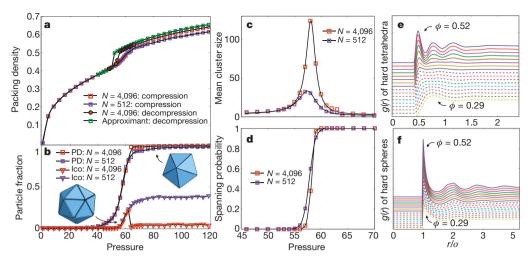


Figure 2 | Thermodynamic and structural properties of the hard tetrahedron fluid. a, Equation of state spanning the transition from the liquid to the solid state. Data are shown for various system sizes. Decompression of both the quasicrystal (N=4,096) and the approximant (unit cell N=82) shows a sharp melting transition. Hysteresis of the compression and decompression curves for the quasicrystal further indicates a first-order transition. For the system with N=512, crystallization is inhibited in many runs, producing a jammed, disordered

glass. **b**, Fraction of tetrahedra participating in pentagonal dipyramids ('PD', right inset) and icosahedra ('Ico', left inset). **c**, Mean cluster size of interpenetrating pentagonal dipyramids. **d**, Spanning probability of the largest cluster of interpenetrating pentagonal dipyramids. **e**, Radial distribution function g(r) of regular tetrahedra for packing fractions ranging from $\phi=0.29$ to $\phi=0.52$. Curves are vertically offset for clarity. **f**, Radial distribution function for the same densities as in **e** for a hard sphere system.

in systems with 13,824 and 21,952 particles (Supplementary Figs 3 and 4). The diffraction pattern obtained by positioning scatterers at the centres of tetrahedra shows rings of Bragg peaks, indicating the presence of long-range order with 12-fold symmetry not compatible with periodicity. Perfect quasicrystals are aperiodic while extending to infinity; they therefore cannot be realized in experiments or simulations, which

are, by necessity, finite. The observed tilings and diffraction patterns with 12-fold symmetry are sufficient in practice for the identification of our self-assembled structures as dodecagonal quasicrystals. Such an identification is in agreement with previous theoretical analysis of random square-triangle tilings²⁶ and findings of dodecagonal quasicrystals in recent experiments^{7–10} and simulations^{27,28}.

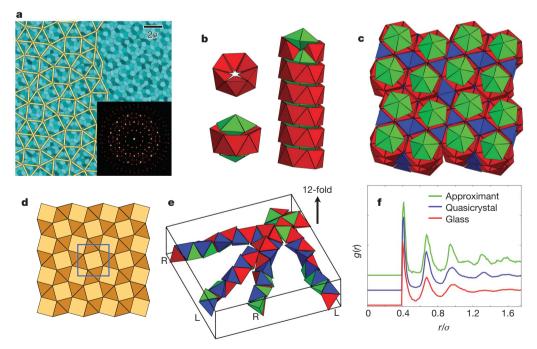


Figure 3 | Structural characterization of the hard tetrahedra dodecagonal quasicrystal and its approximant. a, Network of connected tetrahedra in a slice through a quasicrystal with 8,000 tetrahedra, viewed along the direction of the 12-fold axis. Lines connecting the centre of mass of nearest-neighbour tetrahedra form turquoise pentagons, which correspond to the pentagonal dipyramid network. The tiling structure is highlighted in yellow. A diffraction pattern indicating 12-fold symmetry is shown in the inset. b, The vertices of the tiling are formed by logs comprised of rings of twelve tetrahedra, with neighbouring rings enclosing a pentagonal dipyramid. The packing fraction within the logs can be increased by a tilt of the rings with

respect to the log axis. This allows neighbouring pentagonal dipyramids to avoid each other, as indicated in the figure, where two tetrahedra have been removed from the top ring to expose pentagonal dipyramids. **c**, **d**, Structure (**c**) and tiling (**d**) of the (3,4,3²,4) approximant to the dodecagonal quasicrystal of tetrahedra. Colours are described in the main text. **e**, Similar to the quasicrystal, interpenetrating tetrahelices are present throughout the approximant. Their chirality alternates between left (L) and right (R) by 30° rotations. **f**, Radial distribution functions for the approximant (N = 82), quasicrystal (N = 8,000), and glass (N = 8,000). Curves are vertically offset for clarity.

Quasicrystal approximants are periodic crystals with local tiling structure identical to that in the quasicrystal²⁵. Because they are closely related, and are often observed in experiments, we consider them as candidates for dense packings. The dodecagonal approximant with the smallest unit cell (space group $P\overline{4}n2$) has 82 tetrahedra (Fig. 3c) and corresponds to one of the Archimedean tilings²⁹. At each vertex we see the logs of twelve-member rings (shown in red) capped by single pentagonal dipyramids (green). The logs pack well into squares and triangles with additional, intermediary tetrahedra (blue). The vertex configuration of the tiling is $(3,4,3^2,4)$, as shown in Fig. 3d. Interpenetrating tetrahelices can also be seen in the approximant (Fig. 3e). 'Building' and numerically compressing a unit cell of this ideal structure achieves a packing fraction of $\phi = 0.8479$. If we compress a $2 \times 2 \times 2$ unit cell, the packing fraction marginally increases to $\phi = 0.8503$, the densest packing of tetrahedra yet reported (Supplementary Figs 5 and 6). Compressing approximants with more complex unit cells, more faithful to an ideal quasicrystal, does not further improve the packing (Supplementary Fig. 7), which suggests that a (3,4,3²,4) crystal is the thermodynamically preferred phase at higher pressures.

The fact that the dodecagonal quasicrystal routinely forms in isochoric Monte Carlo simulations of fluids at packing fractions $\phi \geq 0.5$ indicates that the quasicrystal is thermodynamically favoured over the fluid at intermediate pressures. Whether it is stable or metastable relative to the approximant at these pressures is unclear, because the higher entropy of the quasicrystal competes with the higher density of the $(3,4,3^2,4)$ approximant (Fig. 2a) to minimize the Gibbs free energy, and entropically stabilized quasicrystals are known to exist 30,31 . Nonetheless, because the transformation to an approximant is a very slow process 26 , the dodecagonal quasicrystal might be 'practically' stable, even if it is not the thermodynamically stable phase.

Why should square-triangle tilings be preferred for dense packings of tetrahedra? First, we compare the packing fraction of the square tile (22 tetrahedra) to that of the triangle tile (9.5 tetrahedra). Their ratio $\phi_{\rm triangle}/\phi_{\rm square}=19/11\sqrt{3}\approx 0.9972$ is nearly unity, which suggests that tetrahedra pack equally well in both tiles. Second, we note that rings comprising the logs are tilted (Fig. 3b and Supplementary Fig. 8) and the layers of the structure are corrugated (Fig. 1C). This is a direct consequence of the face-to-face packing of tetrahedra where neighbouring logs kiss. As a result, the square tile has a negative Gaussian curvature whereas the triangle tile has a positive one. Alternating the two tiles produces a net zero curvature in the layers, as observed in the quasicrystal and its approximant.

As shown in Fig. 3f, the local structures of the (3,4,3²,4) approximant, the dodecagonal quasicrystal and the disordered glass (as characterized by their radial distribution functions) are very similar. The peak positions are identical: only the peak heights differ. This implies that the local structure of the glass and quasicrystal are only subtly different, and more sensitive measures of local order, as in Fig. 2b, are required. The crucial step during crystallization is the transformation of the percolating pentagonal dipyramid network into layers, and the elimination of icosahedra. This intriguing process will be investigated in subsequent studies.

In conclusion, we report the highest-known packing fraction of regular tetrahedra and show unexpected ways in which they can pack more densely than previously proposed, including the first quasicrystal formed from non-spherical particles. The spontaneous formation of a quasicrystal of hard particles demonstrates that shape alone can produce remarkable structural complexity through solely entropic interactions.

METHODS SUMMARY

We use isobaric and isochoric Monte Carlo simulations with periodic boundary conditions to study systems of N regular tetrahedra, with N ranging from 512 to 21,952. A full Monte Carlo cycle consists of N+1 trial moves including translation plus rotation of a tetrahedron or rescaling of the orthorhombic box. Maximum step sizes are updated occasionally to keep the acceptance probabilities at 30%. Simulations are initialized at low packing fraction in a random configuration and

subsequently compressed to higher densities. The dodecagonal quasicrystals shown in Figs 1 and 3 and Supplementary Figs 1, 3 and 4 are obtained in isochoric simulations at packing fraction $\phi = 0.5$. Crystallization proceeds in three steps: (1) equilibration of the dense, metastable fluid (for example, N = 8,000: $<12 \times 10^6$ Monte Carlo cycles); (2) nucleation and growth $(12-23 \times 10^6$ Monte Carlo cycles); and (3) healing of defects ($>23 \times 10^6$ Monte Carlo cycles). The equation of state in Fig. 2a is calculated by increasing (or decreasing) the external pressure step-wise for compression (or decompression). Longer simulations facilitate equilibration in the transition region. For detecting pentagonal dipyramids and icosahedra in Fig. 2b, nearest neighbours are sampled with a distance cut-off of 0.65σ . The resulting motifs are further screened by projecting the directions of the tetrahedra onto the surface of the unit sphere, and indexing the resulting pattern using spherical harmonics and comparison with an ideal pattern (a pentagon for a pentagonal dipyramid and a dodecahedron for an icosahedron). For $P^* > 120$, compression with conventional Monte Carlo is inefficient. Therefore we apply an alternative method to reach pressures as large as $P^* \ge 10^6$ and obtain maximumdensity packings. The method relies on allowing a small number (of the order of 0.1% of all particles) of minor overlaps (interpenetration of tetrahedra) during box rescaling. All overlaps are subsequently eliminated with isochoric Monte Carlo. Details of our algorithms and the particle data are given in the online-only Methods. Note added in proof: A new result by Kallus et al. 32 indicates a dimer crystal with a packing fraction of $100/117 \approx 0.8547$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 July; accepted 3 November 2009.

- Gardner, M. The Colossal Book of Mathematics: Classic Puzzles, Paradoxes, and Problems 135 (Norton, 2001).
- Betke, U. & Henk, M. Densest lattice packings of 3-polytopes. Comput. Geom. 16, 157–186 (2000).
- 3. Donev, A., Stillinger, F. H., Chaikin, P. M. & Torquato, S. Unusually dense crystal packing of ellipsoids. *Phys. Rev. Lett.* **92**, 255506 (2004).
- Chen, E. R. A dense packing of regular tetrahedra. Discrete Comput. Geom. 40, 214–240 (2008).
- Torquato, S. & Jiao, Y. Dense packings of the Platonic and Archimedean solids. Nature 460, 876–879 (2009).
- 6. Torquato, S. & Jiao, Y. Dense packings of polyhedra: Platonic and Archimedean solids. *Phys. Rev. E* **80**, 041104 (2009).
- Zeng, X. et al. Supramolecular dendritic liquid quasicrystals. Nature 428, 157–160 (2004).
- Hayashida, K., Dotera, T., Takano, A. & Matsushita, Y. Polymeric quasicrystal: mesoscopic quasicrystalline tiling in ABC star polymers. *Phys. Rev. Lett.* 98, 195502 (2007).
- Talapin, D. V. et al. Quasicrystalline order in self-assembled binary nanoparticle superlattices. Nature 461, 964–967 (2009).
- Conrad, M., Krumeich, F. & Harbrecht, B. A dodecagonal quasicrystalline chalcogenide. Angew. Chem. Int. Ed. 37, 1384–1386 (1998).
- Hales, T. C. Historical overview of the Kepler conjecture. Discrete Comput. Geom. 36, 5–20 (2006).
- 12. Hales, T. C. A proof of the Kepler conjecture. Ann. Math. 162, 1065–1185 (2005).
- 13. Glotzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature Mater.* **6**, 567–572 (2007).
- Tang, Z. Y., Zhang, Z. L., Wang, Y., Glotzer, S. C. & Kotov, N. A. Spontaneous selfassembly of CdTe nanocrystals into free-floating sheets. *Science* 314, 274–278 (2006).
- Onsager, L. The effect of shape on the interaction of colloidal particles. Ann. NY Acad. Sci. 51, 627–659 (1949).
- Kirkwood, J. E. in *Phase Transformations in Solids* (eds Smoluchowski, R., Mayer, J. E. & Weyl, W. A.) 67 (Wiley, 1951).
- Bolhuis, P. & Frenkel, D. J. Tracing the phase boundaries of hard spherocylinders. J. Chem. Phys. 106, 666–687 (1997).
- Camp, P. J. & Allen, M. P. Phase diagram of the hard biaxial ellipsoid fluid. *J. Chem. Phys.* 106, 6681–6688 (1997).
- Veerman, J. A. C. & Frenkel, D. Phase-behavior of disk-like hard-core mesogens. Phys. Rev. A 45, 5632–5648 (1992).
- John, B. S., Juhlin, C. & Escobedo, F. A. Phase behavior in colloidal hard perfect tetragonal parallelepipeds. *J. Chem. Phys.* 128, 044909 (2009).
 Conway, J. H. & Torquato, S. Packing, tiling and covering with tetrahedra. *Proc.*
- Natl Acad. Sci. USA 103, 10612–10617 (2006). 22. Kolafa, J. & Nezbeda, I. The hard tetrahedron fluid: a model for the structure of
- water. Mol. Phys. 84, 421–434 (1994).

 23. Frank, F. C. & Kasper, J. S. Complex alloy structures regarded as sphere packings.
- 1. Definitions and basic principles. Acta Crystallogr. 11, 184–190 (1958).
- Fel, L. G. Tetrahedral symmetry in nematic liquid crystals. Phys. Rev. E 52, 702–717 (1995).
- Yamamoto, A. Crystallography of quasiperiodic crystals. Acta Crystallogr. A 52, 509–560 (1996).
- Oxborrow, M. & Henley, C. L. Random square-triangle tilings—a model for twelvefold-symmetrical quasi-crystals. Phys. Rev. B 48, 6966–6998 (1993).

- Roth, J. & Denton, A. R. Solid-phase structures of the Dzugutov pair potential. Phys. Rev. E 61, 6845–6857 (2000).
- 28. Keys, A. S. & Glotzer, S. C. How do quasicrystals grow? *Phys. Rev. Lett.* **99**, 235503 (2007).
- 29. Mikhael, J., Roth, J., Helden, L. & Bechinger, C. Archimedean-like tiling on decagonal quasicrystalline surfaces. *Nature* **454**, 501–504 (2008).
- 30. Steurer, W. Structural phase transitions from and to the quasicrystalline state. *Acta Crystallogr. A* **61**, 28–38 (2005).
- 31. Engel, M. & Trebin, H.-R. Self-assembly of complex crystals and quasicrystals with a double-well interaction potential. *Phys. Rev. Lett.* **98**, 225505 (2007).
- 32. Kallus, Y., Elser, V. & Gravel, S. A dense periodic packing of tetrahedra with a small repeating unit. Preprint at $\frac{htp://arxiv.org/abs/0910.5226}{2009}$.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Air Force Office of Scientific Research supported A.H.-A., P.P.-M. and S.C.G. The National Science Foundation supported A.S.K., A.H.-A. and S.C.G. in the shape-matching analyses that identified local motifs. M.E. was supported by a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft.

Author Contributions A.H.-A. and M.E. performed all simulations and contributed equally to the study. M.E. solved the quasicrystal and approximant structures. A.S.K. performed shape-matching analysis. X.Z., P.P.-M., and R.G.P. proposed and constructed geometric packings. All authors discussed and analysed the results, and contributed to the scientific process. S.C.G., A.H.-A., and M.E. wrote most of the paper, and all authors contributed to refinement of the manuscript. S.C.G. and P.P.-M. conceived and designed the study, and S.C.G. directed the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.C.G. (sglotzer@umich.edu).

doi:10.1038/nature08641 nature

METHODS

We define an ideal, regular tetrahedron as the convex hull of its four vertices $v_1=(1,1,1),\ v_2=(1,-1,-1),\ v_3=(-1,1,-1),\ v_4=(-1,-1,1).$ The edge length of the tetrahedron is $\sigma=\sqrt{8}$ and its volume $V_t=8/3$. The position and orientation of an arbitrary tetrahedron is given by (x,y,z,a,b,c,d), where (x,y,z) is the translation vector and (a,b,c,d) a quaternion describing the rotation. The relation between the quaternion and the rotation matrix is:

$$R = \begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & a^2 - b^2 - c^2 + d^2 \end{pmatrix}$$

Overlap detection. Individual Monte Carlo moves consist of small rotation steps and translation steps with translation vectors chosen randomly in a cubic box of edge length Δr . The most time-consuming part of our algorithm is the overlap check. Tetrahedra are sorted into a cell list with cell size $2\sqrt{3}$, which is the distance beyond which two tetrahedra cannot overlap. Two different overlap detection algorithms are used. They have been designed, written, and tested independently by different co-authors of the present work. Comparing the results of the algorithms on a test set with overlapping and non-overlapping tetrahedra allows independent verification of the codes. The numerical precision of the overlap detection algorithm and therefore the reported packing densities is standard double floating-point precision.

The first algorithm is based on the observation that two convex polyhedra overlap, if and only if at least one edge of one polyhedron intersects one face of the other polyhedron. This means that maximally $2 \times 6 \times 4 = 48$ intersections of lines and triangles have to be evaluated. The algorithm can be optimized by sorting the vertices of each tetrahedron based on their distance from the centre of the other tetrahedron. Only the three edges among the three closest vertices for each tetrahedron require inspection. Additionally, the face that does not include the nearest of the vertices can be discarded. This reduces the necessary checks to a maximum of 18

The second, independent algorithm takes advantage of the fact that two convex polyhedra do not intersect if and only if a plane can be found that completely separates them. In other words, the vertices of one tetrahedron must lie on one side of the plane and the vertices of the other tetrahedron on the other side. Taking two vertices of one tetrahedron and one vertex of the other tetrahedron defines a trial candidate for a separating plane. It can be shown that the study of all such trial candidates is sufficient. Thus, $6\times2\times2=24$ trial candidates need to be checked. Similar to the first algorithm, sorting the vertices allows reduction in the number of candidates. For production simulation runs we use the second overlap detection algorithm only because it is more efficient than the first algorithm.

Compression algorithm. To obtain the maximum density for a given configuration, we quickly compress to very high densities a system first equilibrated using standard isobaric or isochoric Monte Carlo. Such rapid compression is inefficient with the standard isobaric Monte Carlo scheme, so we use a modified Monte Carlo scheme. Our modified scheme is only used to obtain high-density

results; the conventional isochoric and isobaric Monte Carlo methods are used to produce equation-of-state data and to produce the quasicrystals and jammed structures from the fluid.

In the conventional isobaric Monte Carlo algorithm for hard particles, trial volume changes are performed by rescaling the box dimensions. If such a volume change creates an overlap it is discarded, otherwise it is accepted according to the Metropolis criterion for the isobaric ensemble. As the density increases, trial compression moves generate overlaps with larger and larger probability, especially for big systems, making conventional isobaric Monte Carlo slow in equilibrating high densities. To compress our system more efficiently, we introduce a modified scheme that allows a few minor overlaps during compression; these overlaps are then removed to obtain the final configuration.

For our modified scheme, we always accept volume changes, even if they create overlaps. To ensure that the number and amount of overlaps remains small, we use a separate criterion to decide whether the box should be expanded or compressed. We keep track of the acceptance probability p of Monte Carlo translation moves and compare this to a target acceptance probability $p_0 = 0.3$. If $p < p_0$, we apply a compression move, otherwise we apply an expansion move. For our system, we find that rescaling of the box dimensions by a random factor between 1 and $1 + 0.002\Delta r$ for expansion and a random factor between $1 - 0.002\Delta r$ and 1 for compression gives a good balance between fast compression and creating overlaps that are small enough to remove later. The average fraction of particles that overlap during compression is very small, on the order of 0.1% of all particles.

The control parameter for our 'high-density' compression algorithm is the maximum distance of Monte Carlo translation moves, Δr . Although it is not possible to directly measure or control the pressure in our method, we observe that a lower value of Δr corresponds to a higher pressure. For maximum compression, Δr is lowered exponentially to zero.

Our method is extremely simple, fast and robust. By running a short isochoric simulation at the end, we rapidly remove any pre-existing overlaps. This involves implementing the standard translation and rotation Monte Carlo moves, which guarantees that no new overlaps are created and allows sufficient motion, even at very high densities. Using the two independent overlap detection algorithms described above, we ensure that all overlaps are removed for the data reported in this Letter.

Performance. On a single central processing unit (CPU) core, the run time for our most efficient overlap detection scheme is about 5 μ s per particle per Monte Carlo cycle at an intermediate packing density. Typical compression runs for the 13,824 tetrahedra system take a few hours on a single AMD Opteron CPU core with 2.3-GHz clock speed. To obtain good candidates with maximum densities we extended the compression time to a few days. Because of these finite compression times, we restrict the packing densities given in the text to four significant digits. Nucleating the quasicrystal from the fluid in an isochoric simulation and then compressing the quasicrystal with 13,824 particles over 40 million Monte Carlo cycles using our compression algorithm to achieve the data shown in Fig. 1B–D took about one month on a single CPU core.

Catastrophic flood of the Mediterranean after the Messinian salinity crisis

D. Garcia-Castellanos¹, F. Estrada², I. Jiménez-Munt¹, C. Gorini^{3,4}, M. Fernàndez¹, J. Vergés¹ & R. De Vicente¹

The Mediterranean Sea became disconnected from the world's oceans and mostly desiccated by evaporation about 5.6 million vears ago during the Messinian salinity crisis¹⁻³. The Atlantic waters found a way through the present Gibraltar Strait and rapidly refilled the Mediterranean 5.33 million years ago in an event known as the Zanclean flood⁴. The nature, abruptness and evolution of this flood remain poorly constrained⁴⁻⁶. Borehole and seismic data show incisions over 250 m deep on both sides of the Gibraltar Strait that have previously been attributed to fluvial erosion during the desiccation^{4,7}. Here we show the continuity of this 200-km-long channel across the strait and explain its morphology as the result of erosion by the flooding waters, adopting an incision model validated in mountain rivers. This model in turn allows us to estimate the duration of the flood. Although the available data are limited, our findings suggest that the feedback between water flow and incision in the early stages of flooding imply discharges of about $10^8 \,\mathrm{m}^3 \,\mathrm{s}^{-1}$ (three orders of magnitude larger than the present Amazon River) and incision rates above 0.4 m per day. Although the flood started at low water discharges that may have lasted for up to several thousand years, our results suggest that 90 per cent of the water was transferred in a short period ranging from a few months to two years. This extremely abrupt flood may have involved peak rates of sea level rise in the Mediterranean of more than ten metres per day.

The main evidence for a kilometre-scale sea level drop in the Mediterranean is the excavation of canyons by the rivers flowing to the empty sea during the Messinian stage, up to 2,500 m deep in the Nile Delta⁸ and about 1,000 m deep at the mouth of the Rhone⁹. The salt accumulation in the deeper parts of the basin and the deposition of cyclic alternations between brackish and fresh-water sediment of the Lago Mare facies, combined with high-resolution biostratigraphy and astronomically-calibrated magnetostratigraphy^{2,3}, indicate that total disconnection between both sides of the Betic–Rifean orogen started about 5.6 million years ago.

The Messinian salinity crisis finished 5.33 million years ago³, when the Atlantic waters found a way through the present Gibraltar Strait and refilled the Mediterranean in an event known as the Zanclean or post-Messinian flood⁴. There is agreement that this was triggered primarily by tectonic subsidence at the Gibraltar sill, probably related to the sinking of a lithospheric slab under the Betic–Rifean orogen¹⁰, and perhaps in combination with sill erosion¹¹ and sea-level rise. Outburst floods triggered by overspilling of large lakes have induced dramatic changes in surface hydrology and topography in regions as diverse as the Pleistocene Lake Bonneville¹², the Tertiary Ebro basin¹³ (northeast Iberia), or the English Channel¹⁴, but the case of the post-Messinian flood is special because of the enormous size of both the source and the sink basins. The equilibrium level of the isolated Mediterranean during desiccation was between 1,500 m and

2,700 m below present sea level^{6,15}, implying that the flooding water volume was three orders of magnitude larger than that at Lake Bonneville. Because they were based on an arbitrary evolution for the depth of the Gibraltar Strait during the flood, previous estimates of the flood duration yielded divergent values ranging between ten years⁴ and a few thousand years^{5,6}. To quantify and understand the abruptness of the post-Messinian flood we needed to incorporate the dynamics of rock incision as the mechanism that progressively excavated the floodway and let ever increasing flow of Atlantic waters into the Mediterranean basin.

The present maximum depth of the Gibraltar Strait ranges between 284 m at the present Camarinal sill (the shallowest pass between the Atlantic and the Mediterranean; Fig. 1) and about 900 m at the Strait itself. Its present morphology might be affected

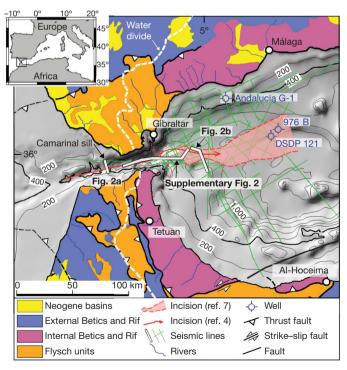


Figure 1 | Geological map and bathymetry of the Gibraltar Arc region. The extent of the erosion channel is shown (after refs 4 and 7). The incision channel cuts 70 km beyond the drainage divide, which we interpret as the result of westwards retrogressive erosion during the post-Messinian flood. The interpretation of the seismic lines is correlated with the three located wells. Fault tectonic deformation has been minor since the Messinian. The water divide between the Atlantic and Mediterranean rivers is shown as a white deshed line.

¹Institut de Ciències de la Terra Jaume Almera, CSIC, Solé i Sabarís s/n, Barcelona, Spain. ²Institut de Ciències del Mar, CSIC, Passg. Marítim Barceloneta, 37-49, Barcelona, Spain. ³Université Pierre et Marie Curie (UPMC) Paris 06, ⁴CNRS, UMR 7193, ISTEP, F-75005, Paris, France.

by the strong streams between both oceanic domains and by tectonic vertical motions after flooding, rather than being an intact relict of Messinian or Zanclean incision. However, the streams did not impede deposition in the strait after the Messinian (Fig. 2a and Supplementary Fig. 2), and therefore they cannot be responsible for the bulk of the present bathymetry. As for tectonic motions, if present at all after the Messinian, they are limited to long-wavelength isostatic or dynamic motions such as those controlling the onset¹⁶ and the end¹⁰ of the Mediterranean isolation, because local fault deformation is minor^{11,17} (see, for example, Fig. 2b).

The strongest evidence for a deep incision channel across the Gibraltar Strait comes from recent boreholes and from seismic data. Drilling cores related to the Africa–Europe tunnel project (Fig. 2a) show a thickness of at least 250 m of flysch breccia redeposited or slumped into a trough carved across the original flysch units (Late Cretaceous to Neogene in age) outcropping in the Iberian and Moroccan sides of the strait 18,19. A similar eastward-oriented incision is observed further to the east, in the Alboran side of the strait⁷ (Fig. 1). Both features have previously been interpreted as subaerial (fluvial) erosion during the Messinian desiccation 4,11,18. In Fig. 2b and Supplementary Fig. 2 we present two sample seismic profiles correlated with Ocean Drilling Program (ODP) site 976 and the commercial well Andalucia G-120,21 through a large set of other publicly available seismic surveys (Fig. 1). These profiles provide evidence for the geometry of this incision and its continuity across the Gibraltar Strait along at least 200 km. As previously recognized18, this erosive channel is incised into Miocene deposits and filled by Pliocene-Quaternary sediments, and merges laterally with the basin-wide Messinian Erosional Surface (MES). In areas unaffected by the widespread

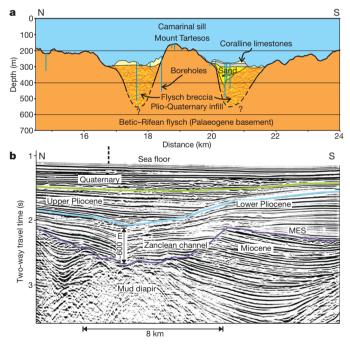


Figure 2 | Evidence for an erosion channel across the Gibraltar Strait.

a, Section across the Gibraltar Strait based on borehole exploration¹⁸. Flysch breccia coming from the Betic–Rifean flysch units fill an erosive trough more than 250 m deep. Whether Mount Tartesos is an autochthonous relict of the resistant flysch¹⁸, or a block slumped from the undermined banks of the flooding channel⁴ is not clear. Vertical exaggeration is 5:1. b, Multichannel seismic profile of Conrad 828 (refs 29, 30) interpreted in this work, based on correlation with the wells located in Fig. 1. The profile shows the dimensions and U-shape of the late-Messinian incision channel in the eastern side of the Gibraltar Strait related to the Zanclean flood. The asymmetry of the channel is partly due to the obliquity of the profile in its northern end (see location in Fig. 1), and partly due to differential isostatic subsidence after the flood¹⁰. MES, Messinian Erosional Surface. Approximate vertical exaggeration is 4:1.

mud diapirism, the geometry of the incision has a U-shaped cross section with a size varying from 650 m depth by 11 km width near the strait to less than 300 m depth per 6 km in different branches of the channel further to the east. The size of this channel is not comparable to any other Messinian palaeovalley observed in the Alboran Sea, but only to canyons carved during the Messinian desiccation by the largest rivers in the Mediterranean^{8,9}. However, the U-shape of this incision (Fig. 2b) and its presence at both the eastern and western sides of the drainage divide (Fig. 1) cast doubt on its formation by subaerial fluvial erosion (typically producing V-shaped valleys) by an eastwardflowing river. Such mechanism would require a large catchment area during the Messinian, but the scarcity of tectonic deformation since that age11,17 suggests that the drainage divide shown in Fig. 1 has not undergone major changes. Recently, U-shaped erosion channels found in the English Channel have been attributed to a megaflood sourced in a large glacial lake in the North Sea¹⁴.

We therefore postulate that the erosion channel observed in Gibraltar (Fig. 1) was excavated by the Zanclean flood. To validate this hypothesis, we calculate the timing of water flow and incision produced during the overspill of the Atlantic basin into the Mediterranean basin by combining a model of rock incision by water with hydrodynamic equations (see Methods). To calculate the flood evolution displayed in Fig. 3, we searched for combinations of the erosional parameters k_b and a that fit a final sill incision of 240 m (a mean value of observed incision in the eastern and western sides of the strait, averaged across the channel). All model runs show a long first period of very little incision owing to the reduced amount of water discharge allowed by the shallow sill depth of 1 m prescribed at the initial time (t=0). As the Gibraltar gate is excavated growing deeper and wider, water flow and incision rate increase exponentially. This situation persists until water flow becomes limited by the rising level of the Western Mediterranean. This event is labelled as stage 1 in Fig. 3. Later, the reduction of the hydrological gradient between the Atlantic Ocean and the western Mediterranean results in reduction of flow velocity, water discharge and incision rate. As the Sicily sill is reached (stage 2), the level of the western basin (and the hydrological gradient) remains constant and the flooding water discharge is transferred to the eastern basin until this is also filled up to that level (stage 3). Afterwards, the whole Mediterranean rises synchronously while the level difference between basins, the water discharge, and the velocity decrease gradually to zero (stage 4).

For comparison, the three model evolutions in Fig. 3 are shown using a time relative to the instant when the western basin reaches the Sicily sill (t_2 , stage 2), which roughly coincides with the time when the rate of sea level rise becomes maximum. Though the exponent of the erosional law a strongly influences t2 (Supplementary Table 1 and Supplementary Fig. 3), the abruptness of the flood remains relatively insensitive to a. This is shown by defining the bulk duration of the flood Δt_b as the time taken by 90% of the total water transfer. High-a, low- k_b model runs imply slow incision at the first stages and therefore a long period of water supply and basin level rise before the catastrophic flow (large t_2). A priori, this could result in a significant refill of the Mediterranean and a reduction in hydraulic gradient, diminishing the abruptness of discharge, but the results show that this occurs only for unrealistic values of the exponent a > 3. Within a values derived for river incision studies, Δt_b changes only from 510 days to 790 days. A complete model parameterization is available in Supplementary Fig. 3.

The amount of incision expected during the post-Messinian flood on the basis of river incision studies (see note 2 in Supplementary Table 1) is comparable in depth and width to the erosion channel observed at the Gibraltar Strait. Using this geometry as a model constraint implies that the Zanclean flood was a catastrophic event (Fig. 3), more abrupt than previously thought^{4–6}, and involved maximum rates of Mediterranean level rise of over 10 m per day. This abruptness has significance not only for its potential effects on ecosystems of the Mediterranean region but also for its palaeoclimatic

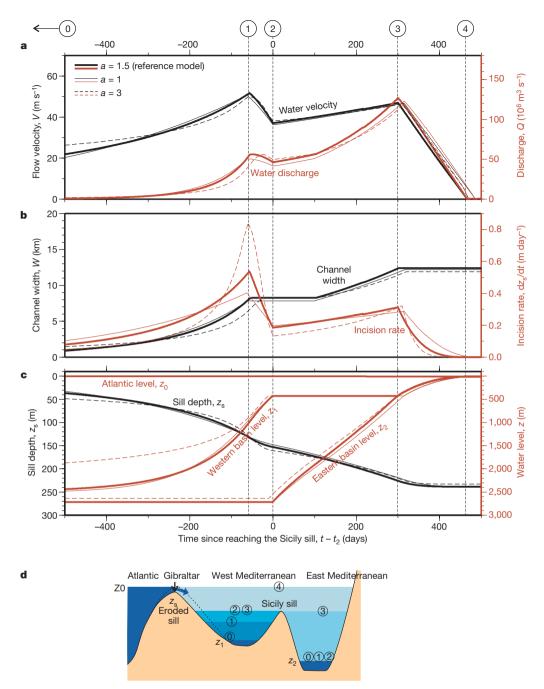


Figure 3 | Evolution of three floods producing a final incision of 240 m, calculated for different exponents of the erosion law a. a, Water velocity and water discharge through the Gibraltar Strait; b, Channel width and incision rate; c, Sill depth (black lines) and level of the Atlantic, the Western Mediterranean, and Eastern Mediterranean (red lines). For comparison purposes, time is relative to the time when the Sicily Sill is reached t_2 (Supplementary Table 1). The three floods start with a sill depth of 1 m.

effects, because a smaller (by two orders of magnitude) outburst flooding at Lake Agassiz has been related to a global cold period around 12,000 years ago 22 . The peak discharge across the Gibraltar Strait reached more than $10^8\,\mathrm{m}^3\,\mathrm{s}^{-1}$ at a speed of over $40\,\mathrm{m}\,\mathrm{s}^{-1}$, only months before flood completion, and produced maximum incision rates exceeding 0.4 m per day. For comparison, the Amazon mean discharge is only $1.5\times10^5\,\mathrm{m}^3\,\mathrm{s}^{-1}$ and the Lake Missoula late glacial catastrophic flood has been estimated in $10^7\,\mathrm{m}^3\,\mathrm{s}^{-1}$ (ref. 23). The Messinian flood implied a dissipation of gravitational potential energy of about $1.6\times10^{22}\,\mathrm{J}$, similar to the heat transport along the Gulf Stream in a year, and $\sim\!4\%$ of the kinetic energy of the K-T Chicxulub meteorite impact 24 .

Although peak discharges take much longer to arrive for large a values, the bulk of the water flow is concentrated in a similar amount of time. Circled numbers refer to the five stages shown in the cartoon \mathbf{d} : (0) initial time; (1) time of maximum incision rate; (2) western Mediterranean level reaches the Sicily Sill (t_2); (3) eastern basin level reaches the Sicily Sill; and (4) The Mediterranean becomes full.

These estimates are consistent with the exceptionally rapid restoration of deep marine conditions recorded at the Messinian–Pliocene boundary²⁵. High-resolution sedimentological studies of this boundary^{26,27} show a brief freshwater influence on the mineralogy, fauna and stable-isotope composition of carbonates over only 15 cm of sediment in ODP site 975. These might reflect the initial flooding period of relatively slow water flow predicted in our calculations, before stage 1. The flood evolution obtained for high values of the incision law exponent (a=3) undergoes little sea level rise in the Mediterranean for the first few thousand years before the catastrophic flow is triggered. Future studies should determine the spatial distribution of the approximately $500\,\mathrm{km}^3$ of rock eroded at the Gibraltar Strait during the flood climax.

We do not envisage the flood as a waterfall, as is often represented: instead, the geophysical data (Supplementary Fig. 2) suggests a huge ramp, several km wide, descending from the Atlantic to the dry Mediterranean with a slope of 1% to 4%, similar to the slope of the present sea floor eastward from Gibraltar. Erosional retreat caused by the flood shifted the sill 30–80 km westwards from its Messinian location at the Gibraltar Strait and shaped the incision channel and the bulk of the strait morphology as we see them today.

METHODS SUMMARY

To use the incision around the Gibraltar Strait as a constraint for the flood velocity, we developed a one-dimensional model that accounts for the feedback between water-flow-controlled incision and sill-depth-controlled water flow. The formulation, based on previous river incision studies²⁸ and on hydrodynamic formulae, is detailed in the Methods. In essence, the model is based on the approach that incision rate dz_s/dt underneath a water flow is a power-law function of basal shear stress τ_b :

$$\frac{\mathrm{d}z_{\mathrm{s}}}{\mathrm{d}t} = k_{\mathrm{b}}(\tau_{\mathrm{b}})^{a} \tag{1}$$

where $k_{\rm b}$ and a are positive constants. An analytical solution of this equation coupled to slope-driven water flow shows that sill incision grows exponentially with time in the early stages of flooding, the speed of this incision being dependent on the lithological erodibility $k_{\rm b}$ and the effective slope on the Mediterranean side of the sill. For the post-Messinian flood, erosion rate doubles in timescales of ten to a hundred years, showing that feedback between incision and water flow is a key control of the timing of the flood.

The interplay between incision (as the floodgate opener) and slope reduction due to the replenishment of the Mediterranean is calculated using an explicit finite-difference time-iterative technique, starting with an initial sill depth of $z_s=1~\mathrm{m}$ at t=0. At each time step, water discharge is calculated based on the depth of the sill and then sill incision is calculated based on basal shear stress and effective slope S (hydrological gradient). As the Mediterranean becomes filled, S gradually decreases to zero. The calculated water discharge is passed from the Atlantic Ocean to the western Mediterranean basin and, if the Sicily sill (430 metres below sea level) is reached, to the eastern Mediterranean basin, accounting for a reconstructed hypsogram of the Messinian Mediterranean (after ref. 6).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 April; accepted 28 September 2009.

- Hsü, K. J., Ryan, W. B. F. & Cita, M. B. Late Miocene desiccation of the Mediterranean. Nature 242, 240–244 (1973).
- Clauzon, G., Suc, J.-P., Gautier, F., Berger, A. & Loutre, M.-F. Alternate interpretation of the Messinian salinity crisis: controversy resolved? *Geology* 24, 363–366 (1996).
- Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J. & Wilson, D. S. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400, 652–655 (1999).
- Blanc, P.-L. The opening of the Plio-Quaternary Gibraltar Strait: assessing the size of a cataclysm. Geodin. Acta 15, 303–317 (2002).
- Hsü, K. J. M. B. C. i. t. a. W. B. F. & Ryan. The origin of the Mediterranean environments. *Init. Rep. Deep Sea Drilling Project*, 13, 1203–1235 (US Government Printing Office, 1973).
- Meijer, P., Th. & Krijgsman, W. A quantitative analysis of the desiccation and refilling of the Mediterranean during the Messinian Salinity Crisis. Earth Planet. Sci. Lett. 240, 510–520, doi:10.1016/j.epsl.2005.09.029 (2005).
- Campillo, A., Maldonado, A. & Mauffret, A. Stratigraphic and tectonic evolution of the western Alboran sea: Late Miocene to recent. Geo-Mar. Lett. 12, 165–172 (1992)
- Barber, P. M. Messinian subaerial erosion of the proto-Nile Delta. Mar. Geol. 44, 1981 253–272 (1981).
- Clauzon, G. The Messinian Var canyon (Provence, southern France): paleogeographic implications. Mar. Geol. 27, 231–246 (1978).
- Govers, R. Choking the Mediterranean to dehydration: the Messinian salinity crisis. Geology 37, 167–170, doi:10.1130/G25141A.1 (2009).

- Loget, N. & Van Den Driessche, J. On the origin of the Strait of Gibraltar. Sedim. Geol. 188–189, 341–356 (2006).
- 12. O'Connor, J. E. Hydrology, Hydraulics, and Gomorphology of the Bonneville Flood GSA Special Paper 274 1–90 (Geological Society of America, 1993).
- Garcia-Castellanos, D., Vergés, J., Gaspar-Escribano, J. M. & Cloetingh, S. Interplay between tectonics, climate and fluvial transport during the Cenozoic evolution of the Ebro Basin (NE Iberia). J. Geophys. Res. 108 (B7), 2347, doi:10.1029/2002JB002073 (2003).
- 14. Gupta, S., Collier, J. S., Palmer-Felgate, A. & Potter, G. Catastrophic flooding origin of shelf valley systems in the English Channel. *Nature* **448**, 342–345 (2007).
- Blanc, P.-L. Improved modelling of the Messinian Salinity Crisis and conceptual implications. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 238, 349–372, doi:10.1016/ j.palaeo.2006.03.033 (2006).
- Duggen, S. & Hoernle, K. v. d. Bogaard, P., Rüpke, L. & Phipps-Morgan, J. Deep roots of the Messinian salinity crisis. *Nature* 422, 602–606 (2003).
- Iribarren, L., Vergés, J., Camurri, F., Fullea, J. & Fernandez, M. The structure of the Atlantic-Mediterranean transition zone from the Alboran Sea to the Horseshoe Abyssal Plain (Iberia-Africa plate boundary). *Mar. Geol.* 243, 97–119, doi: 10.1016/j.margeo.2007.05.011 (2007).
- Esteras, M. Izquierdo, J. Sandoval, N. G. & Bahmad, A. Evolución morfológica y estratigrafía Plio-Cuaternaria del Umbral de Camarinal (Estrecho de Gibraltar) basada en sondeos marinos. Rev. Soc. Geol. Esp. 13, 539–550 (2000).
- Pliego, J. M. The Gibraltar Strait tunnel. An overview of the study process. Tunnelling Underground Space Technol. 20, 558–569 (2005).
- Comas, M. C. et al. Volume 161. Proc. ODP Init. Rep. 161, doi:10.2973/ odp.proc.ir.161.1996 (1996).
- Tandon, K., Lorenzo, J. M. & de La Linde Rubio, J. Timing of rifting in the Alboran Sea basin—correlation of borehole (ODP Leg 161 and Andalucia A-1) to seismic reflection data: implications for basin formation. *Mar. Geol.* 144, 275–294 (1998).
- 22. Broecker, W. Was the Younger Dryas triggered by a flood? *Science* **312**, 1146–1148 (2006).
- 23. O'Connor, J. E. & Baker, V. R. Magnitudes and implications of peak discharges from glacial Lake Missoula. *Geol. Soc. Am. Bull.* **104**, 267–279 (1992).
- Covey, C., Thompson, S. L., Weissman, P. R. & MacCracken, M. C. Global climatic effects of atmospheric dust from an asteroid comet impact on Earth. *Glob. Planet. Change* 9, 263–273 (1994).
- Rouchy, J. M. & Caruso, A. The Messinian salinity crisis in the Mediterranean basin: a reassessment of the data and an integrated scenario. *Sedim. Geol.* 188–189, 35–67, doi:10.1016/j.sedgeo.2006.02.005 (2006).
- Iaccarino, S. M. & Bossio, A. Paleoenvironment of uppermost Messinian sequences in the Western Mediterranean (site 974, 975, and 978). Proc. ODP Sci. Res. 161, 529–541 (1999).
- Pierre, C., Caruso, A., Blanc-Valleron, M. M., Rouchy, J. M. & Orszag-Sperber, F. Reconstruction of the paleoenvironmental changes around the Miocene-Pliocene boundary along a West-East transect across the Mediterranean. Sedim. Geol. 188–189, 319–340 (2006).
- 28. Whipple, K. X. & Tucker, G. E. Dynamics of the stream-power river incision model; implications for height limits of mountain ranges, landscape response timescales, and research needs. *J. Geophys. Res. B* **104**, 17661–17674, doi:10.1029/1999JB900120 (1999).
- 29. Watts, A. B., Platt, J. P. & Buhl, P. Tectonic evolution of the Alboran Basin. Basin Res. 5, 153–177 (1993).
- 30. Docherty, C. & Banda, E. in *The Tertiary Basins of Spain* (eds Friend, P. & Dabrio, C.) 392–398 (Cambridge University Press (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Meijer for providing the reconstructed Messinian hypsometry of the Mediterranean. Thoughtful and constructive reviews by P. Carling and P. Gibbard, and critical comments from I. Cacho, P. Meijer, A. Camerlenghi, R. Carbonell and D. Brown helped in improving earlier versions of the manuscript. This is a Group of Dynamics of the Lithosphere contribution funded by the Spanish Government through the projects 'topoAtlas' (CGL2006-05493), 'TopoMed' and 'Conturiber'.

Author Contributions D.G.-C. planned the study, performed the modelling and wrote the paper; F.E. and C.G. managed and interpreted the seismic lines; I.J.-M. helped with interpretation and writing; M.F. helped in study design; J.V. and R.D.V. took care of the tectonic aspects. All authors discussed the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.G.-C. (danielgc@ictja.csic.es).

doi:10.1038/nature08555

METHODS

Consider a sill at an average depth $z_s > 0$ (a positive value of depth means below initial ocean level) acting as a water gate between a source basin (the Atlantic Ocean) at level z_0 and a sink basin (the western Mediterranean basin) at level z_1 ($z_1 > z_s > z_0$). Symbols are depicted in Supplementary Fig. 1. The incision rate dz_s/dt under a water flow is generally approached as a power-law function of basal shear stress τ_b (equation (1)). The unit stream power approach, including water velocity V as multiplying factor of τ_b in equation (1), has also been tested, but the predicted floods are more abrupt than those shown here because incision is more concentrated at the fastest flooding stages. River incision studies show that a ranges between 1 and 3 (ref. 28). For a = 1, k_b ranges between 10^{-5} and 2×10^{-4} m yr⁻¹ Pa⁻¹ (ref. 31) and $\sim 10^{-7}$ m yr⁻¹ Pa⁻¹ (ref. 32) for river bed incision, and is 18-40 m yr⁻¹ Pa⁻¹ for unconsolidated soil erosion³³. For a = 1.5, a_b has been estimated at $a \times 10^{-6}$ m yr⁻¹ Pa^{-1.5} (ref. 34).

Shear stress at the sill can be approached as the product of water density ρ , the acceleration of gravity g, the mean water depth of the channel $(z_s - z_0)$ and the slope of the water surface S (also known as hydraulic gradient):

$$\tau_{\rm b} = \rho g(z_{\rm s} - z_0) S \tag{2}$$

We assume that S=H/L, where $H=z_1-z_0$ is the head loss, and length L=100 km, which maximizes the half-width of the Betic–Rifean orogen. For a conservative estimation of S we impose a limit of H<1,000 m (representative for the present depth of the Alboran Sea). We note that adopting higher slopes would result in a more abrupt flood.

To calculate the water flow over the sill and the level of the Mediterranean basins, we use an empirical relationship relating water flow speed V with the hydraulic gradient S (Manning's formula), frequently used to estimate outburst flood discharges¹⁴:

$$V = -\frac{1}{n}R_{\rm h}^{\frac{2}{3}}S^{\frac{1}{2}} \tag{3}$$

where V is the average velocity (in metres per second), n = 0.05 is the roughness coefficient, and R_h is the hydraulic radius (in metres) of the strait connecting the Atlantic and the Mediterranean. The hydraulic radius is a measure of the flow efficiency of a river channel, and because channel width is significantly larger than channel depth, it can be estimated as $R_h \approx z_s - z_0$. River discharge (in cubic metres per second) can be calculated as $Q = W(z_s - z_0)V$, where the W is the width of the channel expressed in metres. At each iteration we update the value of W using a relationship derived from river channel studies³⁵:

$$W = k_{\rm w} Q^{a_{\rm w}} \tag{4}$$

where $a_{\rm w}=0.5$ is an empirically determined constant (see, for example, refs 34 and 35) and $k_{\rm w}=1.2$ is a value comparable to normal rivers that has been calibrated here to account for the final width of the Gibraltar Strait, assuming this coincides with the present strait width $W=14\,{\rm km}$. The model predictions have a very small sensitivity to these parameters, as well as to the assumed initial width.

It is possible to solve the feedback dynamics analytically taking z_0 and S as constant (which is valid as long as head loss is not reduced by the refill of the Mediterranean):

$$\frac{\mathrm{d}z_{\mathrm{s}}(t)}{\mathrm{d}t} = K(z_{\mathrm{s}} - z_{\mathrm{0}})^{a} \tag{5}$$

where $K = k_b (\rho g S)^a$ for K > 0. The solution to equation (5) for a = 1 for the sill depth as a function of time is:

$$z_{s}(t) = z_{s}(0) + ce^{Kt}$$
 (6)

Therefore the sill is incised exponentially with time in the early stages of water flow, and the speed of this growth is dependent mostly on the lithological erodibility $k_{\rm b}$ and the slope in the Mediterranean side S. For the post-Messinian flood, $K=10^{-2}$ to 10^{-1} per year, indicating that erosion rate doubles in timescales of ten to a hundred years and that the feedback between incision and water flow is relevant to the timescales of the post-Messinian flood.

To study a more general scenario incorporating both the role of incision (as the mechanism excavating the water gate) and the head-loss reduction due to the replenishment of the Mediterranean, we numerically solved equations (1) to (4) using an explicit finite-difference time-iterative technique. A time step of 0.1 days is used, starting with an initial sill depth of $z_s = 1$ m below the initial ocean level, taken as $z_0 = 0$. We note that changing the initial sill depth from 1 m to 0.1 m induces a strong delay in the reference flood (t_2 increases from 14 to 47 years; see note 4 in Supplementary Table 1), while the predicted maximum flooding rates undergo otherwise insignificant changes (the flood evolution is just shifted in time). Our model cannot determine whether this initial sill depth is related to tectonic subsidence at the Gibraltar Strait¹⁰, or to global sea level rise, or to erosion of the sill¹¹. For the Mediterranean basins we adopt initial levels of $z_1 = 2,500 \text{ m}$ (west) and $z_2 = 2,700$ m (east) below sea level³⁶. The predicted timing of the flood does not vary substantially for a more conservative initial level of 1,500 m. Global sea level drops 9.5 m as a result of the flood, although this result uses the present global ocean hypsometry as a proxy for the one at Messinian times.

The initial geometry adopted for the flooding channel is conservative in the sense that it is chosen to find a maximum estimate for the duration of the flood. For this reason, we have adopted a low value for both the initial slope (S=1%) and a mean incision (240 m for the examples in Fig. 3). Similarly, we have neglected other mechanisms that may have increased incision during the flood, such as cavitation of the initial slope (S=1%) as a cavitation like it is also implicitly assumed that the observed incision is due to a single flood. If an earlier flood took place, its incision across the sill should have been raised above sea level to close the Mediterranean and, in order to affect our estimation of the amount of incision, should be brought below sea level again before the next flood occurred. In other words, multiple flooding could only induce an overestimation of the amount of incision in the presence of post-flood uplift and desiccation-related subsidence at the sill. These vertical sill motions are exactly the opposite of those predicted for the Messinian choking of the Mediterranean 10. It is therefore unlikely that the incision resulted from multiple flooding. Supplementary Fig. 3a shows the effect of the estimated total incision on the predicted duration of the flood.

- 31. Lavé, J. & Avouac, J. P. Fluvial incision and tectonic uplift across the Himalayas of central Nepal. *J. Geophys. Res.* **106**, 26561–26592, doi:10.1029/2001JB000359 (2001).
- 32. Wobus, C. W., Heimsath, A. M., Whipple, K. X. & Hodges, K. V. Active out-of-sequence thrust faulting in the central Nepalese Himalaya. *Nature* **434**, 1008–1011 (2005).
- Elliot, W. J., Liebenow, A. M., Laflen, J. M. & Kohl, K. D. A compendium of soil erodibility data from WEPP cropland soil field erodibility experiments 1987 & 88. NSERL Report 3 (Ohio State University and USDA Agricultural Research Service, National Soil Erosion Research Laboratory, 1989).
- 34. Attal, M., Tucker, G. E., Whittaker, A. C., Cowie, P. A. & Roberts, G. P. Modeling fluvial incision and transient landscape evolution: Influence of dynamic channel adjustment. *J. Geophys. Res.* 113, F03013, doi:10.1029/2007JF000893 (2008).
- 35. Whittaker, A. C., Cowie, P. A., Attal, M., Tucker, G. E. & Roberts, G. P. Bedrock channel adjustment to tectonic forcing: implications for predicting river incision rates. *Geology* **35**, 103–106, doi:10.1130/G23106A.1 (2007).
- Meijer, P., Th., Slingerland, R. & Wortel, M. J. R. Tectonic control on past circulation of the Mediterranean Sea: a model study of the late Miocene. *Paleoceanography* 19, PA1026, doi:10.1029/2003PA000956 (2004).

Thickness and Clapeyron slope of the post-perovskite boundary

Krystle Catalli¹, Sang-Heon Shim¹ & Vitali Prakapenka²

The thicknesses and Clapeyron slopes of mantle phase boundaries strongly influence the seismic detectability of the boundaries and convection in the mantle. The unusually large positive Clapevron slope found for the boundary between perovskite (Pv) and postperovskite (pPv)1-3 (the 'pPv boundary') would destabilize hightemperature anomalies in the lowermost mantle⁴, in disagreement with the seismic observations⁵. Here we report the thickness of the pPv boundary in $(Mg_{0.91}Fe_{0.09}^{2+})SiO_3$ and $(Mg_{0.9}Fe_{0.1}^{3+})(Al_{0.1}Si_{0.9})O_3$ as determined in a laser-heated diamond-anvil cell under in situ high-pressure (up to 145 GPa), high-temperature (up to $3,000\,\mathrm{K}$) conditions. The measured Clapeyron slope is consistent with the D'' discontinuity⁶. In both systems, however, the pPv boundary thickness increases to $400-600 \pm 100 \,\mathrm{km}$, which is substantially greater than the thickness of the D'' discontinuity $(<30 \text{ km})^{7}$. Although the Fe²⁺ buffering effect of ferropericlase⁸⁻¹⁰ could decrease the pPv boundary thickness, the boundary may remain thick in a pyrolitic composition because of the effects of Al and the rapid temperature increase in the D" layer. The pPv boundary would be particularly thick in regions with an elevated Al content and/or a low Mg/Si ratio, reducing the effects of the large positive Clapevron slope on the buoyancy of thermal anomalies and stabilizing compositional heterogeneities in the lowermost mantle. If the pPv transition is the source of the D'' discontinuity, regions with sharp discontinuities may require distinct compositions, such as a higher Mg/Si ratio or a lower Al content.

The lowest 200–400 km of the mantle (the D'' layer) is known to have seismic properties distinct from those of the overlying mantle, including a laterally varying discontinuity at the top of the layer (the D'' discontinuity)^{11,12}. The discovery of a phase transition in the dominant mantle silicate (the pPv transition)^{1,13,14} has provided new opportunities to understand seismological observations better^{2,15} and to constrain important geophysical parameters^{16,17} in the D'' layer.

Any transition in a material with variable composition should have a finite-depth interval in which both low- and high-pressure (P) phases coexist (we call this depth interval the boundary thickness). For seismic detection, the boundary thickness should be sufficiently small. A thickness of less than 30 km was estimated for the D'' discontinuity. Together with its Clapeyron slope (dP/dT; T, temperature), the thickness of the pPv boundary is critical to understanding mantle convection in the D'' region. The thickness of the pPv boundary has not been well constrained. One thickness and the Clapeyron slope of the pPv boundary in ($Mg_{0.9}Fe_{0.1}^{2+}$) (

After being compressed directly without heating, an amorphized (Mg_{0.91}Fe_{0.09})SiO₃ pyroxene sample was heated for a total of 1.5 h at

1,500–2,700 K and pressures higher than 130 GPa. Under these conditions, the synthesis of a Pv+pPv mixture was observed and the phase assemblage remained stable throughout the heating (Fig. 1a). In the next heating run, at a slightly higher pressure, the sample transformed completely to pure pPv (Fig. 1b). A separate sample was heated to 2,000–3,000 K at 137 GPa, under which conditions we observed the synthesis of pure pPv within 10 min. When the pure pPv sample was decompressed by 5 GPa and heated, Pv diffraction lines appeared in 1 h (Fig. 1c), marking the pPv \rightarrow Pv+pPv transition. These observations indicate that the boundary between Pv+pPv and pPv should be at 133–137 GPa and 2,500–3,000 K. Further decompression of this sample, to 119 GPa, and heating to 2,800 K caused strong growth of Pv diffraction lines. In the other samples compressed to 110–120 GPa, stability of the Pv+pPv mixture was observed during 1.5 h of heating. When the Pv+pPv mixture was decompressed and heated to 2,000 K

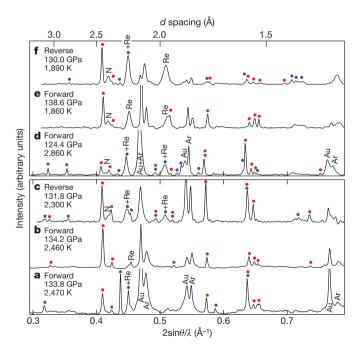


Figure 1 | X-ray diffraction patterns at high pressure and temperature. a, b, c, Pv+pPv (a), $pure\ pPv$ (b) and pPv+Pv (c) in $(Mg_{0.91}Fe_{0.09})SiO_3;$ d, e, f, Pv+pPv (d), $pure\ pPv$ (e) and pPv+Pv (f) in $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3.$ The major diffraction lines are indicated for Pv (blue), pPv (red), Ar (pressure medium), Au (pressure standard), Re (gasket) and N (nitrogen), along with the direction of the transition path (forward or reverse), the pressure and the temperature. Backgrounds were subtracted. The diffraction intensities are plotted as a function of $1/d=2\sin\theta/\lambda$ (d, interplanar distance; θ , diffraction angle; λ , X-ray wavelength).

at 107 GPa, sudden broadening of the pPv lines was observed within 10 min, indicating the instability of pPv.

An amorphous (Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O₃ sample was compressed to 132 GPa, where a Pv+pPv mixture was synthesized during heating. After a total of 55 min of heating at pressures greater than 139 GPa, we observed the transition from a Pv+pPv mixture to pure pPv (Fig. 1e), indicating the stability of the pPv phase. The pure pPv sample was decompressed and heated to 1,800-2,100 K at 130 GPa. After a total of 13 min of heating, a few diagnostic diffraction lines of the Pv phase appeared, indicating the pPv \rightarrow Pv+pPv transition (Fig. 1f). For separate samples, the stability of a Pv+pPv mixture was observed during compression up to 137 GPa at high temperature. Therefore, the boundary between Pv+pPv and pPv should exist between 130 and 140 GPa at 2,000 K. In samples compressed to lower pressures, we observed the stability of a Pv+pPv mixture along both forward and reverse directions between 113 and 137 GPa at 2,000-3,000 K (Fig. 1d). In one of the samples compressed to a lower pressure, synthesis of pure Pv was observed for compression up to 109 GPa at 2,000 K.

The much denser data coverage in $(Mg_{0.91}Fe_{0.09})SiO_3$ tightly constrains the Clapeyron slope of the boundary between Pv+pPv and pPv to $+6.7\pm0.5$ MPa K^{-1} , which is in agreement with the seismological estimation of the D'' discontinuity⁶ and an earlier computational prediction². A larger Clapeyron slope was reported in Mg endmember³, but the discrepancy is probably due to the inconsistency among the different pressure scales used. As differences in pressure can be more reliably determined, the measured thickness should be less affected by this issue.

Our data show that both 9 mol% Fe^{2+} and 10 mol% Fe^{3+} plus 10 mol% Al substantially increase the thickness of the pPv boundary, to 20 ± 5 GPa $(400\pm100\,\mathrm{km})$ and 30 ± 5 GPa $(600\pm100\,\mathrm{km})$, respectively. These values are much greater than the upper bound for the thickness of the D'' discontinuity, $30\,\mathrm{km}$ (ref. 7; Fig. 2). Some sources of uncertainty should be considered. Kinetic effects normally delay a phase transition, resulting in an overestimation of transition pressure along the forward path (from a low-pressure phase to a high-pressure phase) and an underestimation along the reverse path (vice versa). Between the forward and reverse paths there is a 2–5-GPa mismatch in the boundary between Pv+pPv and Pv and we therefore considered the boundary to lie at the mean value. Because the kinetic effects are opposite along these paths, measurement of the Pv boundary along both transition paths reduces errors from these sources.

Heating to sufficiently high temperature for sufficient duration is important in reducing kinetic effects. However, in most previous pPv studies the samples were heated to less than 2,000 K and only a few data exist at higher temperature³. In our study, about one-half of the heating runs were made between 2,500 and 3,000 K. Although each heating run was limited to 15–50 min to prevent overheating of the DAC, heating was repeated at a given pressure for a total heating time of 1–2 h, whereas heating in many previous studies was limited to 3–30 min. As there are steep temperature gradients in laser heating, the colder spots suffer more from kinetic effects. In this study, in order to improve homogeneity in the heated spot, we used Ar as an insulation medium and the sample size was similar to the size of the heating spot.

Our large boundary thicknesses in $(Mg_{0.91}Fe_{0.09})SiO_3$ and $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3$ are consistent with earlier estimations for similar compositions^{19–21} except for those in an *in situ* study¹⁸ in which a negligible pPv boundary thickness in $(Mg,Fe)SiO_3$ was proposed. However, in these studies measurements were not made along both transitional paths and the data coverage was sparse. The *in situ* study¹⁸ was based on diffraction intensity changes in the patterns of Pv+pPv mixtures, which can also be affected by recrystallization and preferred orientation during heating.

In determining the seismic detectability of the pPv transition, it is important to consider the element partitioning among different phases, the shape of the pPv phase-fraction profile in the mixed-phase

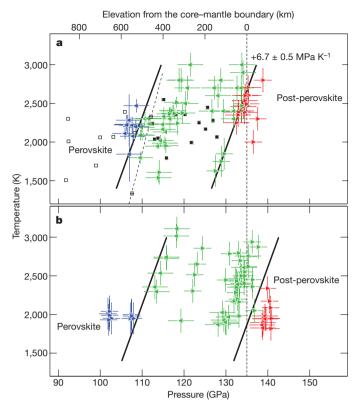


Figure 2 | The pressure-temperature conditions for the stability of Pv, Pv+pPv and pPv. a, Data for $(Mg_{0.91}Fe_{0.09})SiO_3$; b, data for $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3$. The right- and left-pointing triangles represent data points measured along the forward and reverse paths, respectively: Pv, blue; Pv+pPv, green; pPv, red. The solid black lines are the phase boundaries determined from our data, with Clapeyron slope shown in a. The black data points in a are from ref. 18: Pv or Pv+pPv (Pv growth), open squares; Pv+pPv (no growth), crossed squares; Pv+pPv (pPv growth), filled squares. Dashed line, pPv boundary. Error bars, 1σ .

region and the rapid temperature increase in the D $^{\prime\prime}$ layer. To investigate these factors, we calculated the pPv phase-fraction profiles using the ideal-solution model 22 in combination with our measured thicknesses and the results of recent element partitioning studies $^{8-10,23}$ (Fig. 3). Because some discrepancy exists among the measured partition coefficients, we considered all the available data and chose upper bounds for the nonlinear deviations in the pPv phase-fraction profile and the Fe $^{2+}$ buffering effect of ferropericlase (Fp) for presentation in Fig. 3 (Supplementary Information).

Existing partitioning studies are in agreement that Fe^{2+} content follows Fp > pPv > Pv when these phases $coexist^{8-10,19,23}$, and that Fe decreases the pPv transition pressure 10,19 (Supplementary Information). As shown in our calculation along the isotherm at 2,500 K, these factors make the pPv phase-fraction profile nonlinear (Fig. 3d) in such a way that the rate of increase of the pPv phase fraction is greatest near the bottom of the mixed-phase region.

For the measured Fe partition coefficients^{8,9,23}, our calculation indicates that Fe²⁺ buffering by Fp can reduce the pPv boundary thickness to 80–130 km in (Mg,Fe)SiO₃ with 30% Fp. To match the thickness of the D'' discontinuity (<30 km)⁷, the Fp content needs to be greater than the 30% expected for a pyrolitic composition (Fig. 3d).

The buffering effect of Fp also increases the $Pv \rightarrow Pv + pPv$ transition depth, but has very little effect on the $Pv + pPv \rightarrow pPv$ transition depth (Fig. 3d). Because the Fp content increases as the Mg/Si ratio increases, this result suggests that the depth of the pPv transition in (Mg,Fe)SiO₃ increases as the Mg/Si ratio increases. As shown in Fig. 3a, the deepening may result in no intersection between the pPv boundary and the steep geotherm in the D'' layer, and the pPv

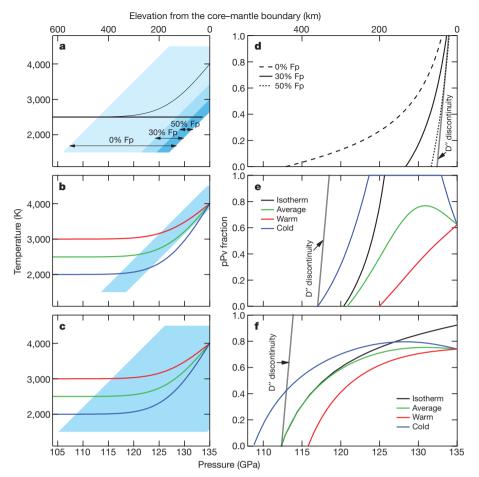


Figure 3 | The Pv+pPv mixed-phase region with mantle geotherms and the pPv phase-fraction profiles in the mixed-phase region. a, d, The ${\rm Fe}^{2^+}$ buffering effect of Fp on the pPv boundary thickness in $({\rm Mg}_{0.91}{\rm Fe}_{0.09}){\rm SiO}_3$. Three cases with different Fp contents (0%, 30% and 50%) are shown, as indicated. In a, the thick and thin lines are the isotherm at 2,500 K and the mantle geotherm, respectively. d shows calculations along the isotherm at 2,500 K. b, e, $({\rm Mg}_{0.91}{\rm Fe}_{0.09}){\rm SiO}_3$ with 30% Fp. The boundary is shifted by -8 GPa to generate the maximum overlap between the pPv boundary and

transition therefore may not exist in regions with an Fp content of 30% or greater.

Because the absolute pressure scale is uncertain ($\pm 5\,\mathrm{GPa}$) over 100 GPa, we do not rule out the possibility of an overlap between the thinned pPv boundary in (Mg,Fe)SiO3 and the geotherm. In Fig. 3b, we intentionally shifted the pPv boundary by $-8\,\mathrm{GPa}$ to produce the maximum overlap. Because the rate of temperature increase in the D'' layer is high and may be comparable to the Clapeyron slope of the pPv boundary, the pPv phase fraction along the mantle geotherm will not increase as rapidly as that along an isotherm (Fig. 3e). In other words, the steep mantle geotherm in the D'' region will suppress the nonlinearity of the pPv phase-fraction profile and prevent completion of the pPv transition.

Studies^{24,25} have found that a significant fraction of iron in Pv and pPv is Fe³⁺ (Fe³⁺/ Σ Fe \approx 60%), through charge-coupled substitution with Al. In the case of Al and Fe³⁺, there is no effect from partitioning with other phases in pyrolite, such as Fp, because Al solubility in Fp is very low and CaSiO₃ perovskite remains nearly pure²⁶. Because Fe³⁺ alone does not increase the thickness (Supplementary Information and Supplementary Fig. 2), we attribute the large thickness found in (Mg_{0.9}Fe³⁺_{0.1})(Al_{0.1}Si_{0.9})O₃ to the effect of Al, which is consistent with previous studies^{20,21}. The result shows that the pPv phase fraction may increase rapidly at the top of the mixed-phase region (Fig. 3f). However, the steep temperature increase in the D'' layer reduces the rate of increase of the pPv phase fraction and prevents the pPv phase transition from completing in the lower mantle (Fig. 3f).

the mantle geotherms (see text). **c, f,** $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3$. The black lines in **e** and **f** are the pPv phase-fraction profiles along the isotherm at 2,500 K. The grey lines in **d, e,** and **f** represent a hypothetical linear phase-fraction profile with the thickness of the D'' discontinuity⁷. The mantle geotherms were calculated using the error function combined with temperature estimation for the region¹⁷. Three different geotherms are shown, representing different temperatures: red, warm; green, average; blue, cold.

Mg silicates in a pyrolitic composition are expected to contain 5–15 mol% of Fe and 7–12 mol% of Al (refs 19, 24, 25, 27), which are similar to the compositions of the samples studied here. Although the pPv boundary thickness in (Mg,Fe)SiO₃ may decrease significantly as a result of Fe²⁺ buffering by Fp, the thickness of the pPv boundary in a pyrolitic composition would remain much greater than the thickness of the D'' discontinuity⁷ because of the large increase in the boundary thickness caused by Al and the large radial temperature gradient in the D'' layer. In other words, the D'' layer may consist of mixed phases of Pv+pPv+Fp instead of pPv+Fp, if the bulk composition of the lowermost mantle is pyrolitic.

The large positive Clapeyron slope of the pPv boundary would increase the density contrast between thermal heterogeneities and bulk mantle, intensifying thermal mantle flow as shown by computer simulations⁴, which is not compatible with seismic observations of the large low-shear-velocity provinces in the lowermost mantle and the stable large-scale mantle plume model⁵. However, the large thickness of the pPv boundary would spread the effect of the Clapeyron slope over a wide depth range, reducing the effects of the pPv boundary on the density contrast and therefore stabilizing high-temperature anomalies in the D'' region. This would be particularly strong in regions with a high Al content and/or a low Mg/Si ratio, influencing the dynamic stability of some chemical heterogeneities in the D'' layer.

A sharp D'' discontinuity has been observed in some regions of the D'' layer 16,17. Therefore, the discrepancy in thickness between the pPv boundary in pyrolite-related compositions and the D'' discontinuity

raises important questions about the origin of the discontinuity. According to our results, the detectability of the pPv boundary could be enhanced in regions with a high Mg/Si ratio and a low Al content, requiring compositional changes for the observation of a sharp pPv boundary. In addition, it has been proposed that strong texturing of pPv might enhance the detectability of the pPv boundary²⁷.

It is notable that multiples of laterally extending reflectors have been identified together with a pair of discontinuities inferred to be the double crossing of the pPv boundary within the bottom 400 km of the lower mantle^{16,17,28}. In addition, strongly tilted sharp boundaries have been documented in seismic studies of the large low-shear-velocity provinces in the lowermost mantle²⁹. These observations are difficult to explain in terms of a single isochemical phase transition.

METHODS SUMMARY

We mixed pyroxene and glass starting materials with 10 wt% gold, which serves as an internal pressure standard. The platelets of these mixtures were loaded in DACs with argon, which serves as a pressure-transmitting and insulating medium. We compressed the samples using bevelled diamond anvils with culets either 75 μm or 100 μm in diameter. In situ high-pressure, high-temperature measurements were conducted in the laser-heated DAC at the GeoSoilEnviroCARS sector of the Advanced Photon Source (Argonne National Laboratory) using double-sided laser heating and an angle-dispersive-diffraction set-up. The data sets for $(Mg_{0.9} Fe_{0.1})(Al_{0.1} Si_{0.9})O_3$ were measured using the same experimental methods, including the pressure scale and pressure medium. This enabled us to measure the thickness and transition pressure among different compositions in an internally consistent fashion.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 August; accepted 19 August 2009.

- Oganov, A. R. & Ono, S. Theoretical and experimental evidence for a postperovskite phase of MgSiO₃ in Earth's D'' layer. Nature 430, 445–448 (2004).
- Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Phase transition in MgSiO₃ perovskite in the Earth's lower mantle. Earth Planet. Sci. Lett. 224, 241–248 (2004).
- 3. Tateno, S., Hirose, K., Sata, N. & Ohishi, Y. Determination of post-perovskite phase transition boundary up to 4400 K and implications for thermal structure in D'' layer. *Earth Planet. Sci. Lett.* **277**, 130–136 (2009).
- 4. Nakagawa, T. & Tackley, P. J. Effects of a perovskite-post perovskite phase change near core-mantle boundary in compressible mantle convection. *Geophys. Res. Lett.* **31**, L16611 (2004).
- Garnero, E. J., Lay, T. & McNamara, A. in Plates, Plumes, and Planetary Processes (eds Foulger, G. R. & Jurdy, D. M.) 79–101 (Geological Society of America, 2007).
- Sidorin, I., Gurnis, M. & Helmberger, D. V. Evidence for a ubiquitous seismic discontinuity at the base of the mantle. Science 286, 1326–1331 (1999).
- Lay, T. Sharpness of the D" discontinuity beneath the Cocos Plate: implications for the perovskite to post-perovskite phase transition. Geophys. Res. Lett. 35, L03304 (2008)
- Kobayashi, Y. et al. Fe-Mg partitioning between (Mg,Fe)SiO₃ post-perovskite, perovskite, and magnesiowüstite in the Earth's lower mantle. Geophys. Res. Lett. 32, L19301 (2005).
- Auzende, A.-L. et al. Element partitioning between magnesium silicate perovskite and ferropericlase: new insights into bulk lower-mantle geochemistry. Earth Planet. Sci. Lett. 269, 164–174 (2008).
- Ono, S. & Oganov, A. R. In situ observations of phase transition between perovskite and CalrO₃-type phase in MgSiO₃ and pyrolitic mantle composition. *Earth Planet. Sci. Lett.* 236, 914–932 (2005).

- Lay, T., Williams, Q. & Garnero, E. J. The core–mantle boundary layer and deep Earth dynamics. *Nature* 392, 461–468 (1998).
- Garnero, E. J. Heterogeneity of the lowermost mantle. Annu. Rev. Earth Planet. Sci. 28, 509–537 (2000).
- 13. Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. Post-perovskite phase transition in MgSiO₃. *Science* **304**, 855–858 (2004).
- Shim, S.-H., Duffy, T. S., Jeanloz, R. & Shen, G. Stability and crystal structure of MgSiO₃ perovskite to the core-mantle boundary. *Geophys. Res. Lett.* 31, L10603 (2004).
- Wookey, J., Stackhouse, S., Kendall, J.-M., Brodholt, J. & Price, G. D. Efficacy of the post-perovskite phase as an explanation for lowermost-mantle seismic properties. *Nature* 438, 1004–1007 (2005).
- Lay, T., Hernlund, J., Garnero, E. J. & Thorne, M. S. A post-perovskite lens and D" heat flux beneath the central Pacific. Science 314, 1272–1276 (2006).
- van der Hilst, R. D. et al. Seismostratigraphy and thermal structure of Earth's coremantle boundary region. Science 315, 1813–1817 (2007).
- Hirose, K., Sinmyo, R., Sata, N. & Ohishi, Y. Determination of post-perovskite phase transition boundary in MgSiO₃ using Au and MgO pressure standards. Geophys. Res. Lett. 33, L01310 (2006).
- Mao, W. L. et al. Ferromagnesian postperovskite silicates in the D" layer of the Earth. Proc. Natl. Acad. Sci. USA 101, 15867–15869 (2004).
- Tateno, S., Hirose, K., Sata, N. & Ohishi, Y. Phase relations in Mg₃Al₂Si₃O₁₂ to 180 GPa: effect of Al on post-perovskite phase transition. *Geophys. Res. Lett.* 32, L15306 (2005).
- Nishio-Hamane, D., Fujino, K., Seto, Y. & Nagai, T. Effect of the incorporation of FeAlO₃ into MgSiO₃ perovskite on the post-perovskite transition. *Geophys. Res.* Lett. 34, L12307 (2007).
- Stixrude, L. Structure and sharpness of phase transitions and mantle discontinuities. J. Geophys. Res. 102, 14835–14852 (1997).
- Sinmyo, R. et al. Partitioning of iron between perovskite/postperovskite and ferropericlase in the lower mantle. J. Geophys. Res. 113, B11204 (2008).
- McCammon, C. Perovskite as a possible sink for ferric iron in the lower mantle. Nature 387, 694–696 (1997).
- 25. Sinmyo, R., Hirose, K., O'Neill, H. S. C. & Okunishi, E. Ferric iron in Al-bearing post-perovskite. *Geophys. Res. Lett.* **33**, L12S13 (2006).
- Hirose, K., Fei, Y., Ma, Y. Z. & Mao, H.-K. The fate of subducted basaltic crust in the Earth's lower mantle. *Nature* 397, 53–56 (1999).
- Murakami, M., Hirose, K., Sata, N. & Ohishi, Y. Post-perovskite phase transition and mineral chemistry in the pyrolitic lowermost mantle. *Geophys. Res. Lett.* 32, L03304 (2005).
- Hutko, A. R., Lay, T., Revenaugh, J. & Garnero, E. J. Anticorrelated seismic velocity anomalies from post-perovskite in the lowermost mantle. Science 320, 1070–1074 (2008).
- Ni, S., Tan, E., Gurnis, M. & Helmberger, D. Sharp sides to the African superplume. Science 296, 1850–1852 (2002).

 $\begin{tabular}{ll} \textbf{Supplementary Information} is linked to the online version of the paper at www.nature.com/nature. \end{tabular}$

Acknowledgements This work is supported by the US National Science Foundation (NSF) grant EAR0738655 (S.-H.S.) and a US Department of Energy (DOE) National Nuclear Security Administration Stewardship Science Graduate Fellowship (K.C.). A. Kubo and B. Grocholski assisted in X-ray measurements. Discussion with T. L. Grove and R. D. van der Hilst improved the paper. This work was performed in the GeoSoilEnviroCARS sector of the Advanced Light Source (APS), Argonne National Laboratory. GeoSoilEnviroCARS is supported by the NSF and the DOE. Use of the APS is supported by the DOE.

Author Contributions K.C. and S.-H.S. prepared and made the measurements on $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3$ and $(Mg_{0.91}Fe_{0.09})SiO_3$, respectively. V.P. assisted in the synchrotron measurements. K.C. and S.-H.S. conducted the data analysis and calculations. S.-H.S. and K.C. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.-H.S. (sangshim@mit.edu).

doi:10.1038/nature08598 nature

METHODS

Starting materials. The starting materials were a natural pyroxene with a composition of $(Mg_{0.9}Fe_{0.0})SiO_3$ and a synthetic glass with a composition of $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3$. The pyroxene sample was also used in our recent X-ray diffraction study of the equation of state and crystal structure of pPv^{30} . The glass starting material was synthesized from a molar mixture of $0.9MgSiO_3+0.05Fe_2O_3+0.05Al_2O_3$ by the containerless method under an O_2 atmosphere to ensure that all Fe remained Fe $^{3+}$ (ref. 31). The purity of the glass starting material was examined by synchrotron X-ray diffraction and Raman spectroscopy. The valence state of Fe in Pv synthesized from the $(Mg_{0.9}Fe_{0.1})(Al_{0.1}Si_{0.9})O_3$ glass was confirmed to be 3+ by synchrotron Mössbauer spectroscopy.

Sample loading. The starting material was powdered and mixed with 8–10 wt% gold for the internal pressure scale. A thin platelet of the sample-plus-gold mixture (thickness, $<5\,\mu m)$ was loaded in either a 35- μm or 50- μm hole in a pre-indented Re gasket (Supplementary Fig. 1). Ar was cryogenically loaded as a pressure and insulation medium. The platelet was supported by grains with the same composition as the starting material, to prevent direct contact with the thermally conductive diamond anvils.

Measurements. At GeoSoilEnviroCARS, a monochromatic X-ray beam (energy, 37 or 40 keV) was focused onto an area of $5 \times 5 \,\mu\text{m}^2$ on the sample and coaxially

aligned with two Nd:YLF laser beams focused on both sides of the sample in the DAC. The size of the laser beam focus ($20\,\mu m$) is comparable to the size of the sample (20– $30\,\mu m$). Diffraction images were measured using the MarCCD detector. The tilt of the charge-coupled-device detector and the sample-to-detector distance were calibrated by measuring the diffraction images of CeO₂. The diffraction images were integrated into one-dimensional patterns using the Fit2D software³². The unit-cell volume of gold, constrained by 2–5 diffraction lines, was used, in combination with its equation of state³³, to calculate the pressure. The temperature of the sample was estimated by fitting the thermal radiation from the sample to Planck's equation³⁴.

- 30. Shim, S.-H. *et al.* Crystal structure and thermoelastic properties of (Mg_{0.91}Fe_{0.09})SiO₃ postperovskite up to 135 GPa and 2700 K. *Proc. Natl. Acad. Sci. USA* 105, 7382–7386 (2008).
- Tangeman, J. A. et al. Vitreous forsterite (Mg₂SiO₄): synthesis, structure, and thermochemistry. Geophys. Res. Lett. 28, 2517–2520 (2001).
- 32. Hammersley, A. P. *Fit2d: An Introduction and Overview.* ESRF Internal Report (European Synchrotron Radiation Facility, 1997).
- 33. Tsuchiya, T. First-principles prediction of the *P–V–T* equation of state of gold and the 660-km discontinuity in Earth's mantle. *J. Geophys. Res.* **108**, 2462 (2003).
- 34. Jeanloz, R. & Heinz, D. L. Experiments at high temperature and pressure: laser heating through the diamond cell. *J. Phys. (Paris)* **45**, C8–83–C8-92 (1984).

Reproductive skew and selection on female ornamentation in social species

Dustin R. Rubenstein^{1,2,3} & Irby J. Lovette³

Male animals are typically more elaborately ornamented than females¹. Classic sexual selection theory notes that because sperm are cheaper to produce than eggs², and because males generally compete more intensely for reproductive opportunities and invest less in parental care than females³, males can obtain greater fitness benefits from mating multiply^{2,4}. Therefore, sexual selection typically results in male-biased sex differences in secondary sexual characters^{1,4}. This generality has recently been questioned, because in cooperatively breeding vertebrates, the strength of selection on traits used in intrasexual competition for access to mates (sexual selection^{1,4}) or other resources linked to reproduction (social selection^{5,6}) is similar in males and females^{7,8}. Because selection is acting with comparable intensity in both sexes in cooperatively breeding species, the degree of sexual dimorphism in traits used in intrasexual competition should be reduced in cooperative breeders⁶. Here we use the socially diverse African starlings (Sturnidae) to demonstrate that the degree of sexual dimorphism in plumage and body size is reduced in cooperatively breeding species as a result of increased selection on females for traits that increase access to reproductive opportunities, other resources, or higher social status. In cooperative breeders such as these, where there is unequal sharing of reproduction (reproductive skew) among females, and where female dominance rank influences access to mates and other resources, intrasexual competition among females may be intense⁷ and ultimately select for female trait elaboration9. Selection is thereby acting with different intensities on males and females in cooperatively versus non-cooperatively breeding species, and female-female interactions in group-living vertebrates will have important consequences for the evolution of female morphological, physiological and behavioural traits.

Recent challenges to the utility¹⁰ and ubiquity¹⁰ of Darwin's classic concept of sexual selection¹—which emphasizes selection on elaborate traits in males—have been based on the recognition that similar selection may act on females. Indeed, a growing awareness of the potential for social or intrasexual competition among females^{5,6} has helped spur interest in sexual selection and trait elaboration in females^{11,12}. Yet despite this renewed attention, we do not fully understand why females in some species are highly ornamented, whereas females in related species are drab or dull. Moreover, it is not clear if the same underlying principles and mechanisms that commonly operate in males act similarly in females.

The concept of sexual selection, which argues that elaborate traits may confer a fitness gain in the form of a mating advantage, is often invoked to explain ornamentation in animals. Although formulated upon the idea of differential investment in gamete production and parental care by the sexes^{2,3}, variation in reproductive success resulting from intrasexual competition for reproductive opportunities is also central to sexual selection theory⁹. In species with unequal sharing of reproduction among same-sex individuals—a pattern termed

high reproductive skew—variance in lifetime reproductive success is a direct consequence of intrasexual competition for breeding and will be higher in the sex with the greatest potential rate of reproduction¹³. Since the intensity of sexual selection on traits that enhance breeding opportunities should be directly related to the variance in direct fitness among members of each sex¹⁴, selection should thus act more strongly on the sex with the higher variance in reproductive success. In most species, males have higher variance in reproductive success than do females¹⁵, and the traditional attention¹ given to elaborate, sexually selected male traits reflects this general trend. However, this pattern is often absent or reversed in cooperatively breeding vertebrates that live in family groups, where females frequently have a higher variance in reproductive success than males^{8,16}. This femalebiased reproductive variance in cooperative breeders has been observed in both high skew societies (singular breeders)8 and lower skew societies (plural breeders)16, as well as in both birds and mammals8. Although not all cooperatively breeding species show this relationship, the general pattern emerges because skew is high in both males and females in many cooperatively breeding species, whereas it tends to be high only in males in non-cooperatively breeding species. Indeed, skew is often higher in females than in males in cooperatively breeding species because non-pair-bonded males may have the option to reproduce via extrapair matings (which reduces male skew and variance in reproductive success), whereas non-paired females generally cannot reproduce at all8.

Selection is likely to favour traits that improve a female's reproductive success directly via intrasexual competition for reproductive opportunities (sexual selection⁴), or indirectly via intrasexual competition for increased social status or access to resources other than mates (for example, food, nest sites, territories) linked to reproduction (social selection^{5,6}). In cooperative breeders, sexual or social selection on traits used in intrasexual competition is expected to be relatively strong in both sexes7, and should therefore result in a reduced degree of sexual dimorphism and females that are similar in ornamentation to males, as has been suggested for some gregarious and territorial species⁶. This hypothesis assumes that the same traits influence competitive success in both sexes. However, if males and females use different traits in intrasexual competition, the opposite might be expected and the sexes would be predicted to show divergent characters and an increased degree of sexual dimorphism in cooperatively breeding species¹².

The pattern of reversed reproductive variance in cooperatively breeding vertebrates⁸ provides an unusual opportunity to test the relationship between reproductive skew and the traits that ultimately give rise to sexual dimorphism. To explore whether the degree of sexual dimorphism in traits used in intrasexual competition is reduced in cooperatively breeding species, we analysed patterns of morphological sexual dimorphism in a comparative framework

¹Columbia University, Department of Ecology, Evolution and Environmental Biology, 10th Floor Schermerhorn Extension, 1200 Amsterdam Avenue, New York, New York 10027, USA. ²University of California, Berkeley, Department of Integrative Biology, Museum of Vertebrate Zoology, 3060 Valley Life Sciences Building, Berkeley, California 94720, USA. ³Cornell Laboratory of Ornithology, Fuller Evolutionary Biology Program, 159 Sapsucker Woods Road, Ithaca, New York 14850, USA.

using a complete species-level phylogeny of a socially variable lineage of birds, the 45 species of African starlings (Sturnidae). Nearly 40% of the starling species endemic to Africa are cooperative breeders, making them an appropriate group in which to examine the evolutionary causes and consequences of vertebrate sociality¹⁷. Although few of these starlings have been studied in detail, variance in reproductive success is higher in females than in males in cooperatively breeding superb starlings, *Lamprotornis superbus*¹⁶, as has been found generally in cooperatively breeding vertebrates⁸. Additionally, reproductive skew is high in all cooperatively breeding starlings; although most species are singular breeders, even the few plural breeding species have relatively high skew among females^{17–19}.

Our phylogenetic hypothesis for the African Sturnidae was constructed using Bayesian likelihood methods applied to extensive mitochondrial and nuclear intron sequences, and its high topological resolution provides a robust framework for testing evolutionary hypotheses^{17,20} (Fig. 1a). African starlings are a visually diverse and highly ornamented group that exhibits a range of plumage colours, feather structures and other ornaments ranging from iridescent plumage to crests²¹. We measured sexual dimorphism in two traits likely to be used in intrasexual competition: plumage and body size (wing length) dimorphism. These characters were mapped onto the tree and analysed using phylogenetically controlled comparative analyses.

We found that the degree of sexual dimorphism in both plumage and body size was lower in cooperatively breeding than in non-cooperatively breeding starling species. Cooperative breeders had sexually monomorphic plumage, whereas non-cooperative breeders had sexually dimorphic plumage ($\chi^2 = 14.7$, n = 45, P = 0.0001; phylogenetically controlled: likelihood ratio (LR) = 7.12, P < 0.0001; Fig. 1b). Additionally, females and males had more similar wing lengths in cooperatively breeding species than in non-cooperatively breeding species (t=2.13, d.f.=42, P=0.039; phylogenetically independent contrasts (PICs): $F_{1,10} = 5.11$, P = 0.029; Fig. 1c). Although a variety of other demographic and ecological factors have been shown to influence sexual dimorphism in birds²², these alternative explanations probably do not explain the patterns observed in African starlings, as the relationships we observed among body size, plumage dimorphism and social system were not related to any of a suite of alternative hypotheses (Supplementary Discussion), but instead are likely to reflect selection acting with different intensities on traits associated with intrasexual competition.

Cooperatively breeding starlings show reduced sexual dimorphism compared to non-cooperative breeders, but does this pattern in cooperative breeders stem from reduced selection on male traits, or increased selection on female traits used in intrasexual competition? Although there is no reason to predict that selection acts less intensely on males in cooperatively breeding species, there is reason to predict that the intensity of selection acting on female traits is increased in cooperatively breeding starlings, as it is in other group-living species^{22,23}. For example, increased intrasexual competition for reproductive opportunities is particularly intense in cooperative breeders with high reproductive skew where one or a few females in a group monopolize the reproduction^{7,8}. Similarly, since dominance rank and access to resources like food or breeding sites are important in many cooperative breeders, intrasexual competition for such resources probably explains the masculinized morphological and physiological adaptations in spotted hyenas and some other cooperatively breeding vertebrates^{9,24}.

As quantifying the absolute strength of selection is difficult in a comparative study, we assessed the relative intensities of selection acting on male and female phenotypic traits. We found that in cooperatively breeding starlings with reduced sexual dimorphism, selection is likely to be acting more strongly on female plumage elaboration and body size rather than more weakly on these traits in males. In many avian groups, plumage dimorphism with drab females and ornamented males is the ancestral condition, and transitions to monomorphism are associated with female trait elaboration 11,22,23. We assessed plumage traits in male and female African

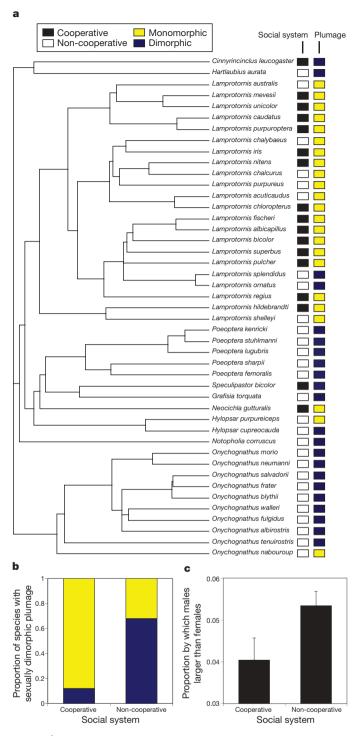


Figure 1 | Molecular phylogeny of African starlings with their associated behavioural and morphological traits. a, The tree is based on combined analysis of mitochondrial and nuclear intron sequences. Social and morphological characters are indicated at each terminal species; a key is given above. b, A greater proportion of a non-cooperative species had sexually dimorphic plumage than did cooperative species, suggesting a reduced degree of plumage dimorphism in cooperative species. c, Males and females in cooperative species had wings that were more similar in length than those in non-cooperative species (mean \pm s.e.), suggesting a reduced degree of size dimorphism in cooperative species.

starlings and found that females were less ornamented than males with dull or drab plumage in all 22 sexually dimorphic species, but equally ornamented as males with similarly bright or elaborate plumage in all 23 sexually monomorphic species. Therefore, females in cooperatively breeding species were more ornamented than those in

non-cooperatively breeding species ($\chi^2 = 23.3$, n = 45, P < 0.0001; phylogenetically controlled: LR = 7.12, P < 0.0001). We also examined the relationship between male and female wing length ($F_{1,42} = 3,579.06$, P < 0.0001, R = 0.99; Fig. 2a) and calculated the residuals to compare the relative intensity of selection on the sexes in cooperative and non-cooperative taxa. We found that males had longer wings than females in all species (Fig. 2a). However, in cooperatively breeding species females tended to have relatively longer wings (positive residuals), whereas in non-cooperatively breeding species, females tended to have relatively shorter wings (negative residuals) (t = 1.87, d.f. = 42, P = 0.034; PICs: $F_{1,10} = 3.82$, P = 0.04; Fig. 2b).

To further test the hypothesis that reduced sexual dimorphism in cooperatively breeding species reflects enhanced selection on female traits^{22,23}, we explored the relative magnitudes of variance in male and female body size. Simple tests of which sex has greater variance are not illuminating, as there are opposing predictions about the relationship between selection intensity and phenotypic trait variance. Because selection on a trait often leads to reduced phenotypic and genetic variance in that trait^{25,26}, the sex with the greater variance could be predicted to be under relatively weaker selection. Alternatively, because traits under strong selection might be expected to have high, not low, genetic variance under some conditions²⁷, the sex with greater phenotypic variance could be expected to be under relatively greater selection. Available evidence from other taxa provides support for both relationships: comparative studies of avian plumage suggest that phenotypic traits under greater sexual selection are often as variable as traits not under sexual selection²⁸, whereas experimental studies in insects suggest that genetic variance in sexually selected traits can be small²⁹ or large³⁰, depending upon the trait. Here, however, the key prediction is that under either scenario, males and females of cooperatively breeding species should have similar variances in body size, whereas these traits in males and females of non-cooperatively breeding species should have dissimilar variances because the relative intensity of selection is more similar between the sexes in the cooperatively breeding species.

Our results support this key prediction of dissimilar variances: we found no significant effect of sex or social system on the variance in wing length, but there was a significant interaction (sex: $F_{1,45.84} = 0.68$, P = 0.42, social system: $F_{1,40.03} = 2.89$, P = 0.097; interaction: $F_{1,41.08} = 7.84$, P = 0.0078) such that the variance in wing length did not differ between males and females in cooperative breeders ($F_{1,42.48} = 1.55$, P = 0.22) but, in non-cooperative breeders, females had greater variance in wing length than males ($F_{1,44.97} = 8.02$, P = 0.0069; Fig. 3a). The

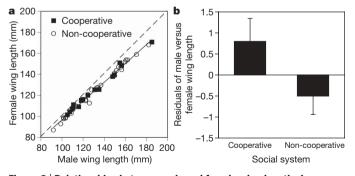
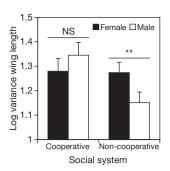


Figure 2 | Relationships between male and female wing lengths in cooperatively and non-cooperatively breeding African starlings. a, Male and female wing lengths were positively related, but males always had longer wings than females. Cooperative and non-cooperative starlings showed the same relationship (interaction: $F_{1,40}=13.97,\,P=0.10$); the dashed line indicates the line of equality. b, The residuals of the phylogenetically controlled relationship between male and female wing lengths were significantly greater in cooperative than in non-cooperative starlings (mean \pm s.e.), suggesting that in cooperative species, females had relatively longer wings, whereas in non-cooperative species, females had relatively shorter wings.



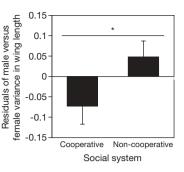


Figure 3 | Patterns of variance in male and female wing lengths in cooperatively and non-cooperatively breeding African starlings. a, In non-cooperative starlings, females had greater variance in wing length than males, but in cooperative starlings males and females had similar levels of variance in wing length (mean \pm s.e.). The double asterisks indicate significance at P < 0.01. b, The residuals of the phylogenetically controlled relationship between male and female variance in wing length were significantly greater in non-cooperative starlings than in cooperative starlings (mean \pm s.e.). The single asterisk indicates significance at P < 0.05.

critical result is that phenotypic variance differs between males and females in non-cooperatively breeding species, but not in cooperatively breeding species. We confirmed this result in a phylogenetic framework by first calculating the residuals of the relationship between male and female variance in wing length and then using a comparative analysis to compare the residuals of this relationship in cooperative and non-cooperative species. We found a significant difference in the residuals of cooperative and non-cooperative breeders (PICS: $F_{1,10} = 6.14$, P = 0.033; Fig. 3b). To further control for within-species comparisons in a phylogenetic framework, we performed an additional series of analyses that yielded similar results to those presented here (Supplementary Data).

Collectively, our results suggest that the patterns of sexual dimorphism related to social system in African starlings derive from intersexual differences in the relative intensity of selection acting on female traits used in intrasexual competition in societies with different levels of reproductive skew. Biologists have long recognized the link between mating systems and sexually selected traits¹³. Here, we describe a similar link between social systems, patterns of reproductive skew and secondary sexual characteristics. Sexual selection theory typically emphasizes intrasexual competition among males^{1,9,12}, yet female-female competition for reproductive opportunities and other resources related to breeding has important fitness consequences in many cooperatively breeding vertebrates⁷, particularly those with high female reproductive skew where females compete for breeding status. When competition among females is intense, selection on traits that improve access to mates, other resources, or higher social status within the group will be strong. We have shown that this increased femalefemale competition in high skew societies results in more ornamented females and therefore less sexual dimorphism. Because high female reproductive skew is common in cooperatively breeding vertebrates⁸, the patterns of reduced sexual dimorphism and increased selection on females in cooperatively breeding starlings are likely to be observed in other social vertebrates that live in groups with unequal sharing of reproduction. Although Darwin largely ignored the importance of female intrasexual competition as a selective force in the evolution of elaborate female traits¹, female–female interactions in group-living vertebrates will have important consequences for not only the evolution of female morphological traits like those reported here, but also for physiological and behavioural traits associated with intrasexual competition9.

METHODS SUMMARY

A detailed description of phylogenetic reconstruction methods, samples and markers was published previously²⁰. Briefly, we used Bayesian Markov chain Monte Carlo (MCMC) methods and maximum parsimony to reconstruct the

phylogeny from a substantial nuclear and mitochondrial DNA sequence character matrix. We included all 45 Sturnidae species that are of African origin and endemic to mainland Africa and its satellite islands 17,20 . Species were classified as 'cooperative' or 'non-cooperative' using published species accounts and behavioural observations 17 . Plumage dimorphism was scored as 'dimorphic' or 'monomorphic' as determined from published descriptions and visual comparison of many specimens of all taxa. Body size dimorphism was calculated as the difference in male and female wing length relative to male wing length, as measured from museum specimens (n=1,614). Phylogenetically controlled comparative analyses were conducted using Pagel's discrete algorithms in Mesquite v2.01 or using the BRUNCH algorithm in CAIC v2.6 (see Supplementary Methods for all program references).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 September; accepted 27 October 2009.

- Darwin, C. The Descent of Man, and Selection in Relation to Sex (Murray, London, 1871).
- 2. Bateman, A. J. Intra-sexual selection in *Drosophila*. Heredity 2, 349–368 (1948).
- 3. Trivers, R. L. in Sexual Selection and the Descent of Man 1871–1971 (ed. Campbell, B.) 136–179 (Aldine, 1972).
- 4. Andersson, M. Sexual Selection (Princeton Univ. Press, 1994).
- West-Eberhard, M. J. Sexual selection, social competition, and evolution. Proc. Am. Phil. Soc. 123, 222–234 (1979).
- West-Eberhard, M. J. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58, 155–183 (1983).
- Clutton-Brock, T. H. et al. Intrasexual competition and sexual selection in cooperative mammals. Nature 444, 1065–1068 (2006).
- Hauber, M. E. & Lacey, E. A. Bateman's principle in cooperatively breeding vertebrates: the effects of non-breeding alloparents on variability in female and male reproductive success. *Integr. Comp. Biol.* 45, 903–914 (2005).
- Clutton-Brock, T. H. Sexual selection in males and females. Science 318, 1882–1885 (2007).
- Roughgarden, J., Oishi, M. & Akcay, E. Reproductive social behavior: cooperative games to replace sexual selection. Science 311, 965–969 (2006).
- Amundsen, T. Why are female birds ornamented? Trends Ecol. Evol. 15, 149–155 (2000).
- 12. Clutton-Brock, T. H. Sexual selection in females. Anim. Behav. 77, 3–11 (2009).
- Clutton-Brock, T. H. in Evolution from Molecules to Men 457–481 (Cambridge Univ. Press, 1983).
- 14. Arnold, S. J. Bateman's principle and the measurement of sexual selection in plants and animals. *Am. Nat.* 144, S1–S22 (1994).
- Clutton-Brock, T. H. Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems (Univ. Chicago Press, 1988).
- Rubenstein, D. R. Temporal but not spatial environmental variation drives adaptive offspring sex allocation in a plural cooperative breeder. Am. Nat. 170, 155–165 (2007).
- Rubenstein, D. R. & Lovette, I. J. Temporal environmental variability drives the evolution of cooperative breeding in birds. Curr. Biol. 17, 1414–1419 (2007).

- Rubenstein, D. R. Female extrapair mate choice in a cooperative breeder: trading sex for help and increasing offspring heterozygosity. Proc. R. Soc. Lond. B 274, 1895–1903 (2007)
- 19. Wilkinson, R. Social organization and communal breeding in the chestnut-bellied starling (*Spreo pulcher*). *Anim. Behav.* **30**, 1118–1128 (1982).
- Lovette, I. J. & Rubenstein, D. R. A comprehensive molecular phylogeny of the starlings (Aves: Sturnidae) and mockingbirds (Aves: Mimidae): congruent mtDNA and nuclear trees for a cosmopolitan avian radiation. *Mol. Phylogenet. Evol.* 44, 1031–1056 (2007).
- 21. Craig, A. J. F. K. & Hartley, A. H. The arrangement and structure of feather melanin granules as a taxonomic character in African starlings (Sturnidae). *Auk* 102, 629–632 (1985).
- 22. Badyaev, A. V. & Hill, G. E. Avian sexual dichromatism in relation to phylogeny and ecology. *Annu. Rev. Ecol. Syst.* **34**, 27–49 (2003).
- 23. Irwin, R. E. The evolution of plumage dichromatism in the New World blackbirds: social selection of female brightness? *Am. Nat.* **144**, 890–907 (1994).
- Dloniak, S. M., French, J. A. & Holekamp, K. E. Rank-related maternal effects of androgens on behaviour in wild spotted hyenas. *Nature* 440, 1190–1193 (2006).
- 25. Bulmer, M. G. The effect of selection on genetic variability. *Am. Nat.* **105**, 201–211 (1971).
- 26. Lande, R. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402–416 (1979).
- Rowe, L. & Houle, D. The lek paradox and the capture of genetic variance by condition dependent traits. Proc. R. Soc. Lond. B 263, 1415–1421 (1996).
- 28. Delhey, K. & Peters, A. Quantifying variability of avian colours: are signaling traits more variable. *PLoS One* **3**, e1689 (2008).
- 29. Van Homrigh, A., Higgie, M., McGuigan, K. & Blows, M. W. The depletion of genetic variance by sexual selection. *Curr. Biol.* 17, 528–532 (2007).
- Kotiaho, J. S., Simmons, L. W. & Tomkins, J. L. Towards a resolution of the lek paradox. *Nature* 410, 684–686 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by grants from NSF to I.J.L. (DEB-0515981) and a postdoctoral fellowship from the Miller Institute for Basic Research in Science at the University of California, Berkeley to D.R.R. We thank E. Lacey, B. Lyon, B. McCleery, D. I. Rubenstein, R. Safran, A. L. Talaba and the Lacey laboratory for advice and assistance. P. Hohohnle helped conduct and interpret the MIPoD analyses. We acknowledge S. Edwards, J. Trimble (Harvard Museum of Comparative Zoology), R. Prum, K. Zyskowski (Yale Peabody Museum), J. Cracraft, P. Sweet (American Museum of Natural History), N. Rice (Academy of Natural Sciences, Philadelphia), M. Braun, D. James (National Museum of Natural History) and R. Prys-Jones (British Museum) for access to their collections.

Author Contributions D.R.R. conceived the project and analysed the data. D.R.R. and I.J.L. jointly collected the data and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.R. (dr2497@columbia.edu).

doi:10.1038/nature08614 nature

METHODS

Phylogeny reconstruction. We included all 45 Sturnidae species that are of African origin and endemic to mainland Africa and its satellite islands^{17,20}. A detailed description of phylogenetic reconstruction methods, samples and markers was published previously²⁰. Briefly, we used Bayesian Markov chain Monte Carlo (MCMC) methods and maximum parsimony to reconstruct the phylogeny from a substantial nuclear and mitochondrial DNA sequence character matrix.

Characters. Species were classified as 'cooperative' or 'non-cooperative' using published species accounts and behavioural observations¹⁷. Plumage in African starlings is highly ornamented, but the type and biochemical structure of the ornaments varies among clades²¹. Because it is unlikely that a single type of ornament is used for inter- or intrasexual communication across species, plumage dimorphism was scored conservatively and categorically as 'dimorphic' or 'monomorphic' as determined from published descriptions and visual comparison of many specimens of all taxa. We then determined whether females had reduced ornamentation compared to males (that is, more drab or dull in colour), or if they were as ornamented as males using published species accounts¹⁷; there were no cases in starlings where females were more ornamented than males.

We calculated body size dimorphism as the difference in male and female wing length relative to male wing length, as measured from museum specimens (n=1,614). Although selection can operate more strongly on plumage dichromatism than body size dimorphism in some avian species, these two measures of sexual dimorphism were significantly related in African starlings, as species with plumage dimorphism also had greater body size dimorphism (t=2.05, d.f.=42, P=0.047). Wing lengths were measured by the same researcher (I.J.L.) to the nearest 0.5 mm. Variance in wing length was log-transformed to meet the assumptions of normality. The median number of males and females measured for each species was 20 individuals of each sex (females: mean = 17.4; range = 0–23 individuals; males: mean = 18.8 males; range = 1–23). Because of the rarity of specimen material, we were able to measure only one female *Pholia femoralis* and one *Onychognathus neumanni*, and hence we did not include these two species in our analyses of continuous traits. For all other species, we measured at least five individuals of each sex.

Statistical analyses. To examine the relationship between plumage dimorphism and sociality, we used a chi-squared test (non-phylogenetically controlled) and Pagel's discrete algorithms in the program Mesquite v2.01 (phylogenetically controlled) (see Supplementary Methods for all program references). This latter analysis uses a continuous-time Markov model and allows for tests of correlated evolution as well as for the order and direction of evolution for binary traits.

Models of evolution are fitted to the data and phylogeny using maximum likelihood, and described by the log-likelihood of the model. Correlated evolution is detected by comparing a model where two traits are allowed to evolve independently with one in which they are constrained to co-evolve. The transition rates for the traits were treated as identical, in a one-parameter model where the forward transition (α) was equal to the backwards transition (β). This test compares the log-likelihoods of the model of independent evolution (H_0) to that of the mode of dependent evolution where the traits are linked (H_1). Monte Carlo simulations of the independent and dependent models were run 10,000 times, and the likelihood ratio (LR) and P-value were reported.

To examine the relationship between body size dimorphism and social system, we used a *t*-test (non-phylogenetically controlled) and CAIC v2.6 to conduct comparative analyses on phylogenetically independent contrasts. We used the BRUNCH algorithm with the binary discrete characters for social system (cooperative versus non-cooperative). CAIC performs a regression on the contrasts and reports an *F*-ratio with the appropriate degrees of freedom for the number of phylogenetically independent contrasts (PICs) calculated from the phylogeny.

To test the prediction that the reduced degree of sexual dimorphism in cooperatively breeding starlings results from increased selection on females, we used a chi-squared test (non-phylogenetically controlled) and Pagel's discrete algorithms in the program Mesquite v2.01 (phylogenetically controlled) to examine the relationship between female ornamentation and social system. We used a correlation analysis to examine the relationship between male and female wing length, followed by a *t*-test of the residuals (non-phylogenetically controlled) and a phylogenetically controlled analysis using the BRUNCH algorithm in CAIC to examine whether males or females differed more in cooperative and non-cooperative species. Because we made an a priori prediction that the reduced sexual dimorphism resulted from increased selection on females, we used one-tailed tests for these analyses.

Generalized linear mixed models were used to examine the relationship between social system, sex and variance in wing length. Sex, social system and their interaction were included in the model as fixed effects, species was included as a random effect to control for comparisons within a species, and wing length was included to control for differences in body size among species. Contrasts were used to examine least square means in post-hoc analyses. Results were confirmed in a phylogenetic framework by comparing the residuals of the relationship between female and male variance in wing length in cooperative and non-cooperative species using the BRUNCH algorithm in CAIC.

Chiral blastomere arrangement dictates zygotic left-right asymmetry pathway in snails

Reiko Kuroda^{1,2,3}, Bunshiro Endo², Masanori Abe² & Miho Shimizu²

Most animals display internal and/or external left-right asymmetry. Several mechanisms for left-right asymmetry determination have been proposed for vertebrates¹⁻¹⁰ and invertebrates^{1,2,4,9,11-14} but they are still not well characterized, particularly at the early developmental stage. The gastropods Lymnaea stagnalis and the closely related Lymnaea peregra have both the sinistral (recessive) and the dextral (dominant) snails within a species and the chirality is hereditary, determined by a single locus that functions maternally¹⁵⁻¹⁸. Intriguingly, the handedness-determining gene(s) and the mechanisms are not yet identified. Here we show that in L. stagnalis, the chiral blastomere arrangement at the eight-cell stage (but not the two- or four-cell stage) determines the left-right asymmetry throughout the developmental programme, and acts upstream of the Nodal signalling pathway. Thus, we could demonstrate that mechanical micromanipulation of the third cleavage chirality (from the four- to the eight-cell stage) leads to reversal of embryonic handedness. These manipulated embryos grew to 'dextralized' sinistral and 'sinistralized' dextral snails—that is, normal healthy fertile organisms with all the usual left-right asymmetries reversed to that encoded by the mothers' genetic information. Moreover, manipulation reversed the embryonic nodal expression patterns. Using backcrossed F7 congenic animals, we could demonstrate a strong genetic linkage between the handedness-determining gene(s) and the chiral cytoskeletal dynamics at the third cleavage that promotes the dominant-type blastomere arrangement. These results establish the crucial importance of the maternally determined blastomere arrangement at the eight-cell stage in dictating zygotic signalling pathways in the organismal chiromorphogenesis. Similar chiral blastomere configuration mechanisms may also operate upstream of the Nodal pathway in left-right patterning of deuterostomes/vertebrates.

Embryonic morphogenesis along the anterior–posterior and dorsalventral axes has been well characterized, but that of the left–right axis has only recently begun to be elucidated^{1–14}. In some vertebrates, directional nodal flow appears to be important for left–right asymmetry determination^{5–8}, and in invertebrates such as *Drosophila* and *Caenorhabditis elegans*, the actin cytoskeleton and an associated type I myosin^{11,12}, and a G α protein regulating spindle orientation¹⁴, seem to be involved, respectively. However, the initial symmetry-breaking steps are not yet known. An intracellular model of early left–right patterning has been highlighted, where asymmetric gene expression is initiated by oriented cytoskeletal elements⁹ and ion flux¹⁰.

We have focused on the snail *L. stagnalis* as a system with unique advantages for studying chiromorphogenesis from the molecular to the organismal level. The correlation of the handedness of the spiral blastomere cleavage with the directions of shell coiling was first proposed based on the observation of sinistral *Physa heterostropha* and dextral *Lymnaea columella*¹⁹. Indeed clockwise and anticlockwise

third cleavage have been observed for the dextral and the sinistral snails within a species of *L. peregra* and *L. stagnalis*, respectively^{17,20}.

To examine directly the role and timing of blastomere arrangements on shell coiling, we have used micromanipulation to reverse the genetically specified third-cleavage directions in both sinistral and dextral embryos of L. stagnalis. At metaphase-anaphase (for dextral snails) or telophase (for sinistral snails) of the third cleavage, we used two glass rods to push the animal surface of each blastomere in the directions opposite to the normal third cleavage (Fig. 1a, d). A constant mechanical force was applied to each cell during furrow ingression until contact between newly formed adjacent pairs of a micromere and a macromere was established (Fig. 1b, e). The manipulation reversed the spindle orientations, shifted the cleavage planes towards the opposite directions, and created the chirality-inverted embryos (Fig. 1c, f and Supplementary Fig. 1). Judging from the blastomere configurations and intercellular contacts, nearly 78% (71/91) of sinistral embryos were successfully reversed to 'dextralized' eight-cell stage embryos and 78% (67/86) of dextral embryos to 'sinistralized' eight-cell stage ones (Supplementary Table 1).

Spiral cleavage is characterized by the alternating clockwise and anticlockwise cleavages during the third to fifth cycles. At the fourth cleavage, after the recurrent blastomere compaction during postmitotic phase in which the morphological chirality of embryos was seemingly lost, both the artificially reversed dextral (48/67) and sinistral (38/71) embryos exhibited rotation in the opposite sense to their mothers' genetic information, keeping the alternative rotation direction in successive cleavages (Supplementary Fig. 1, Supplementary Table 1). The reversed fourth cleavage direction was clearly seen in the fluorescence imaged cell lineage tracing (Fig. 1k-q, with their bright field image in Fig. 1g-j). 72% (48/67) of the dextral embryos and 54% (38/71) of the sinistral embryos displayed totally inverted blastomere arrangements at the 16-cell stage (Supplementary Table 1). Thus, spindle orientation at spiral cleavage stages is controlled by spatial arrangement of blastomeres that is determined by the previous cell cleavage event.

We further investigated whether the effect of totally inverted blastomere arrangements at the third cleavage continues throughout the whole developmental programme of the snail. We incubated the manipulated embryos in a glass capillary tube with their natural capsular fluid. After about 17 days, 31% (13/42) of the inverted dextral embryos developed into juvenile snails, and remarkably, all of them had the sinistrally coiled shells with completely reversed features (Fig. 2a–c, Supplementary Table 1). 46% (10/22) of the reversed sinistral embryos also developed into juvenile snails, and again they all showed the dextrally coiled shells (Fig. 2g–i, Supplementary Table 1). They can be compared with the normal sinistral (Fig. 2m–o) and the normal dextral (Fig. 2s–u) snails. The juvenile snails were reared to adults (Fig. 2d, j; compare their respective normal snails,

¹Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan. ²Kuroda Chiromorphology Team, ERATO-SORST, JST, Komaba, Meguro-ku, Tokyo 153-0041, Japan. ³Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

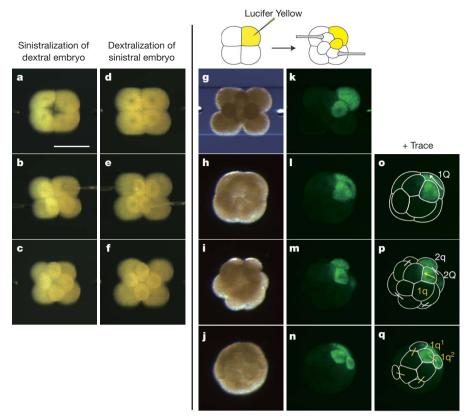


Figure 1 | Reversal of the third cleavage directions by micromanipulation and the resultant 8-, 12- and 16-cell stage embryos. Dextral embryos at the metaphase-anaphase (a) and sinistral embryos at the telophase (d) of the third cleavage were manipulated. The first quartet of micromeres getting generated was continuously pushed towards the direction opposite to normal by glass rods (sinistrally for the dextral embryo (b) and dextrally for the sinistral embryo (e)), which resulted in chirality-inverted sinistral-type (c) and dextral-type (f) eight-cell embryos, respectively. Fluorescenceimaged cell-lineage tracing was carried out by injecting Lucifer Yellow dye into one quadrant of the four-cell stage sinistral embryo, then reversing the

Fig. 2p, v) and their internal organ asymmetry was examined in detail. Fully grown 'sinistralized' and 'dextralized' snails had pulmonary sac, anus, male and female genital pores open at the left or right side of the body (Fig. 2e, k), just like the normal sinistral (Fig. 2q) and the dextral (Fig. 2w) snails, respectively, and internal organs, such as heart, stomach, liver coiling and gut looping, with the shape and positions (Fig. 2f, l) just like the normal sinistral (Fig. 2r) and dextral (Fig. 2x) snails, respectively. Thus, the chirality-reversed embryos at the eightcell stage developed to situs inversus. We did not observe situs solitus or situs ambiguus. The reversed-coiled snails were fertile, and produced sinistral or dextral progenies dictated by their genotype and not the reversed body handedness (Supplementary Table 1).

Although chirality is the most prominent at the third cleavage, it can be traced back to the first and second cleavages²¹. We altered, by manipulation, the directions of blastomere rotations of both the sinistral and dextral embryos at the first or the second cleavage to produce reversed blastomere configuration at the four-cell stage. However, the manipulated embryos all reverted to the original-type third cleavage (Supplementary Fig. 2). We also observed that sinistral embryos occasionally showed dextral-type blastomere arrangement at the four-cell stage even in the egg capsules, but they all showed normal anticlockwise cleavage at the third division. Thus, macromere–micromere cell contacts at the eight-cell stage embryo appear to be the first determining step for asymmetric development of snails.

We have previously reported that dextral and sinistral snail embryos are not mirror images of each other at the third cleavage (refs 20, 22). The dominant dextral snails exhibit spiral deformation (SD) and spindle inclination (SI), while the recessive sinistral snails chirality by manipulating as in **d**–**f** and culturing them. The resultant dextral-type eight-cell stage sinistral embryo (\mathbf{g} , \mathbf{k}) was compacted (\mathbf{h} , \mathbf{l} , \mathbf{o}) and then cleaved into 12- (\mathbf{i} , \mathbf{m} , \mathbf{p}) and 16-cell (\mathbf{j} , \mathbf{n} , \mathbf{q}) embryos, which arose from the typical non-synchronous division of macromeres (1Q) and micromeres (1q). Each blastomere of 1Q (\mathbf{o}) and 1q (\mathbf{p}) divided in the dextral-type anticlockwise direction and produced their descendants 2q-2Q (\mathbf{p}) and 1q¹-1q²(\mathbf{q}), respectively. a– \mathbf{j} , Bright field image; k– \mathbf{q} , fluorescence image with outline of blastomeres (\mathbf{o} – \mathbf{q}). Arrows (\mathbf{o} , \mathbf{p}) indicate the spindle orientation. Scale bar, 100 µm.

do not show them²⁰ (see below). SD is a helical deformation of the blastomeres at the metaphase–anaphase, and SI is a spiral orientation of the four spindles, as a consequence of SD, before the cleavage furrow ingression²⁰. We have succeeded in making F_7 congenic animals, which inherit 99.2% of sinistral strain-derived and 0.8% of the dextral strain-derived genome. Remarkably, SD and SI were observed in all the dextral embryos oviposited by F_7 animals that inherit the dextrality gene(s), but not in any of the sinistral embryos oviposited by F_7 snails devoid of the dextrality gene(s). Thus, the organismal handedness-determining gene(s) is strongly linked to, or is, the gene that induces or activates SD and/or SI. We made dextral snails by pushing the micromeres of sinistral embryos from the telophase without SD.

These results suggest that chiral blastomere configuration is the key factor in handedness determination, which is achieved by SD and SI genetically in the wild, and by micromanipulation in our experiments. The epigenetic manipulation reprograms the left–right asymmetry determination most probably by altering blastomere arrangement around the 3D organizer which is specified at the 24-cell stage²³. In the case of *C. elegans*, it has been reported¹³ that mechanical treatment at the six-cell stage produced chirality-reversed animals, similar to the case of *L. stagnalis*. Although spindle orientation is important in both species, *L. stagnalis* appears to adopt a different chirality determining pathway (see below). We have studied the orthologues of $G\alpha$ and several cell polarity-related proteins (for example, Par6, atypical PKC) for the sinistral and the dextral *L. stagnalis*, but no chirality-dependent difference was observed in their expression (T. Homma, M.S. and R.K., unpublished results).

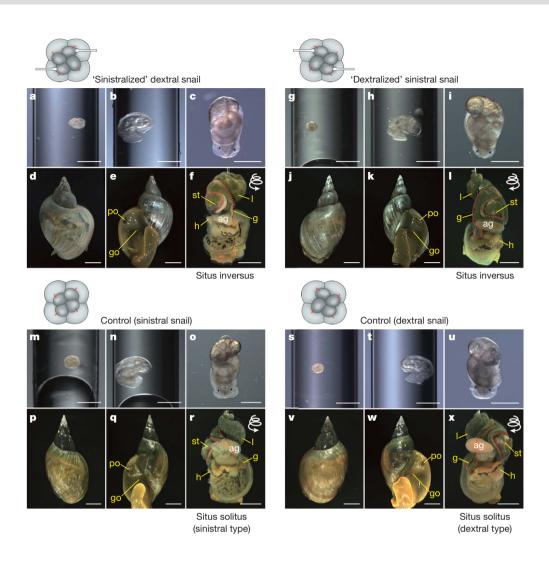


Figure 2 | Chirality-reversed eightcell stage embryos developed to snails with an oppositely-coiled shell and visceral situs inversus. After the third-cleavage manipulations, both sinistralized dextral embryos (a) and dextralized sinistral embryos (g) were raised to adult snails. Non-manipulated sinistral (m) and dextral (s) embryos were also cultured in parallel as control. Development was observed at trochophore (a, g, m, s), veliger (b, h, n, t) and juvenile snail (c, i, o, u) stages. Adult snails were pictured dorsally (d, i, p, v) and ventrally (e, k, q, w). The shell was removed to observe the position of internal organs (f, l, r, x, dorsal view). The sinistralized dextral snails (a-f) and control sinistral snails (m-r) are morphologically identical, so are the dextralized sinistral snails (g-I) and control dextral snails (s-x). ag, albumen gland; g with dotted red line, gut; go, female genital opening; h, heart; l with white coil, liver; st, stomach; po, pulmonary sac opening. Scale bars: **a**-**c**, **g**-**i**, **m**-**o**, **s**-**u**, 0.5 mm; **d**–**f**, **j**–**l**, **p**–**r**, **v**–**x**, 5 mm.

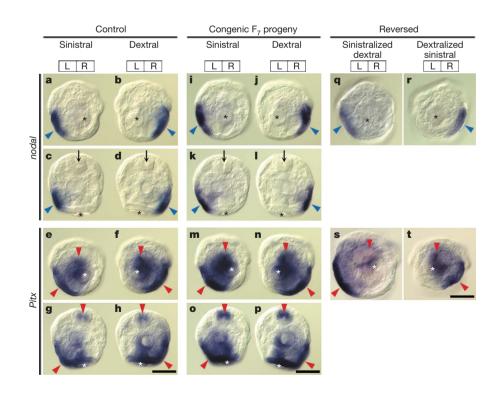


Figure 3 | nodal and Pitx expression in control, congenic F₇ progeny, and chirality-inverted L. stagnalis embryos. L and R indicate left and right sides of late trochophore stage embryos. Pairs of sinistral and dextral embryos were placed side-by-side and pictured in the same field. a, b, e, f, i, j, m, n, q, r, s and t are posterior views, while c, d, g, h, k, l, o and p are dorsal views. Position of the shell gland is marked by an asterisk, and that of the stomodeum by arrows in dorsal view c, d, k and l. nodal (blue arrowhead) is expressed in the left lateral ectoderm, near the left side of the developing shell for the sinistral embryos (a, c), but in the right lateral ectoderm, near the right side of the developing shell for the dextral embryos (**b**, **d**). Pitx (red arrowhead) is expressed in the stomodeum, visceral mass, and left or right side of posterolateral ectoderm for the sinistral (e, g) and dextral (f, h) L. stagnalis embryos, respectively. nodal and Pitx expressions of embryos of F₇ congenic progeny exhibited similar patterns (i–l and m–p) to those of control (parental inbred strain) (a-d and e-h). Chirality reversed embryos, that is, sinistralized dextral (q, s) and dextralized sinistral (r, t) embryos, showed identical expression patterns to the control (a, e) and progeny of F_7 (i, m) sinistral, and the control (b, f) and progeny of F7 (j, n) dextral embryos, respectively. Scale bars, 100 μm.

Asymmetric activation of the Nodal pathway is a conserved feature of deuterostomes for the determination of the asymmetric body plan²⁴. The Nodal pathway does not appear to be involved in Ecdysozoa such as flies and nematodes. However, a recent report²⁵ revealed that the Nodal pathway does operate in gastropods (Lophotrochozoa), as evidenced by the contrasting asymmetric expression of nodal and its downstream Pitx genes in the sinistral snail Biomphalaria glabrata and compared with the dextral snail Lottia gigantea. To determine whether left-right asymmetry at the eight-cell stage affects the nodal expression pattern, we cloned the orthologues of *nodal* and *Pitx* genes and investigated their expression patterns in non-manipulated and manipulated L. stagnalis by whole mount in situ hybridization at the late trochophore stage. For the sinistral L. stagnalis, nodal is expressed in the left lateral ectoderm, near the left side of the developing shell (Fig. 3a, c), whereas it is found in the right lateral ectoderm, near the right side of the developing shell, for the dextral *L. stagnalis* (Fig. 3b, d). Nodal expression was detected first during the 32–64 cell stages (data not shown). Similarly, Pitx is expressed in the stomodeum, and visceral mass, and asymmetrically in the left or right side regions of posterior and lateral ectoderm for the sinistral and dextral L. stagnalis, respectively (Fig. 3e-h), the sinistral case being similar to the results for B. glabrata²⁵. Embryos of progenies of sinistral and dextral F₇ congenic snails exhibited asymmetric *nodal* and *Pitx* patterns (Fig. 3i–p) exactly the same as control strains described above (Fig. 3a-h), indicating that the Nodal pathway acts downstream of the handedness-determining

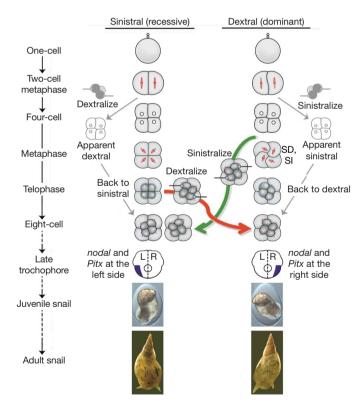


Figure 4 | Determinants of chirality in the snail *L. stagnalis*. This scheme summarizes the development of left- and right-handed snails from the one-cell stage to mature adults. In *L. stagnalis*, reversing the chirality by micromanipulation at the first or second cleavage stage does not alter the organismal chirality, as the manipulated embryos revert to form eight-cell embryos of original handedness (thin arrow). In contrast, embryos whose chirality is reversed by micromanipulation at the third cleavage grow to chirality inverted juvenile and then to healthy and fertile adult snails, with oppositely-coiled shell and situs inversus viscerum (thick arrow). *nodal* and *Pitx* expressions are also reversed by this manipulation. Dextralized snails are produced from sinistral snails without SD (spiral deformation), a unique feature observed only at the third-cleavage metaphase-anaphase of dominant dextral snails, and directly linked to the handedness-determining gene(s).

gene product(s). Remarkably, when chirality at the eight-cell stage was reversed by micromanipulation, the *nodal* and *Pitx* expression patterns at the late trochophore stage were completely reversed (2/2 for *nodal* and 6/6 for *Pitx*), as is clearly seen in Fig. 3q–t. Thus, the maternally-determined blastomere arrangement at the eight-cell stage dictates the zygotic Nodal signalling pathway.

The features of maternal inheritance of chirality in Lymnaea^{15,16} and the correlation of shell coiling handedness with the spiral blastomere cleavage are long established¹⁹. However, the nature of the link between them has remained obscure. In this paper, we show for the first time that the chiral blastomere arrangement at the eight-cell stage, whose cytoskeletal dynamics is directly controlled by the handednessdetermining gene(s), dictates the Nodal pathway at the late trochophore stage, leading to the left-right body asymmetries (summarized in Fig. 4). These results indicate that the role of genetically important SD and SI is to achieve the correct micromere-macromere arrangement of dominantly handed snails at the eight-cell stage. Precisely how a particular blastomere arrangement affects the fate of blastomeres and engagement of the Nodal pathway is unknown, as is the mechanism by which chiral memory is transferred through subsequent spiral cleavage cycles. Studies in *Lymnaea* should provide a tractable system in which to answer such intriguing and fundamental questions that have relevance for left-right asymmetry not only for spiralians but also more widely in other more complex organisms.

METHODS SUMMARY

Our original stocks of both the sinistral and the dextral L. stagnalis were kindly supplied by G. Smit (Vrijie Universiteit) and have been reared in our laboratory over many years, essentially as described earlier¹⁸. Micromanipulation of cleavage directions was performed under a stereomicroscope (Leica) with two handmade glass rods controlled by hydraulic micromanipulators (Narishige) while holding embryos in a groove on an agarose slab. Manipulated and non-manipulated embryos were transferred into glass capillary tubes containing natural capsular fluid and were cultured for 2–3 weeks until they had developed into juvenile snails. Juveniles were transferred to small aquaria and reared to adult. For the fluorescence imaging, four-cell stage embryos, after 5 mM DTT treatment, were injected with Lucifer Yellow by using a conventional hydraulic injection system, and the embryos were observed under a fluoro-stereomicroscope (Leica). For double staining of F-actin and microtubules in embryos, Alexa 488-phalloidin (Molecular Probes) and Cy3-conjugated monoclonal anti-β-tubulin antibody (Sigma) were used. DAPI was used for DNA staining. Images were obtained by fluorescence microscopy (Axio Imager M1, Zeiss). Three-dimensional-reconstruction images were made from z-series of optical sections acquired every 1.0 µm. Whole mount in situ hybridization was performed as described²⁶ except for the following conditions: snail embryos were fixed with 3.2% formaldehyde in MTSTr (50 mM PIPES-KOH, pH 6.9, 25 mM EGTA, 150 mM KCl, 25 mM MgCl $_{\!2}, 0.1\%$ Triton X-100) overnight at 4 °C. We used BM purple (Roche) instead of NBT-BCIP for the detection of signals.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 July; accepted 22 October 2009. Published online 25 November 2009.

- Brown, N. A. & Wolpert, L. The development of handedness in left/right asymmetry. *Development* 109, 1–9 (1990).
- Spéder, P., Petzoldt, A., Suzanne, M. & Noselli, S. Strategies to establish left/right asymmetry in vertebrates and invertebrates. *Curr. Opin. Genet. Dev.* 17, 351–358 (2007).
- Shiratori, H. & Hamada, H. The left-right axis in the mouse: from origin to morphology. *Development* 133, 2095–2104 (2006).
- Vandenberg, L. N. & Levin, M. Perspectives and open problems in the early phases of left-right patterning. Semin. Cell Dev. Biol. 20, 456–463 (2009).
- Nonaka, S., Shiratori, H., Saijoh, Y. & Hamada, H. Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature* 418, 96–99 (2002)
- Okada, Y., Takeda, S., Tanaka, Y., Belmonte, J. C. & Hirokawa, N. Mechanism of nodal flow: a conserved symmetry breaking event in left-right axis determination. *Cell* 121, 633–644 (2005).
- Nonaka, S. et al. De novo formation of left-right asymmetry by posterior tilt of nodal cilia. PLoS Biol. 3, e268 (2005).
- Hirokawa, N., Tanaka, Y., Okada, Y. & Takeda, S. Nodal flow and the generation of left-right asymmetry. *Cell* 125, 33–45 (2006).

- Levin, M. & Palmer, A. R. Left-right patterning from the inside out: widespread evidence for intracellular control. *Bioessays* 29, 271–287 (2007).
- Levin, M., Thorlin, T., Robinson, K. R., Nogi, T. & Mercola, M. Asymmetries in H⁺/K⁺-ATPase and cell membrane potentials comprise a very early step in left-right patterning. *Cell* 111, 77–89 (2002).
- Spéder, P., Ádám, G. & Noselli, S. Type ID unconventional myosin controls leftright asymmetry in *Drosophila*. *Nature* 440, 803–807 (2006).
- Hozumi, S. et al. An unconventional myosin in *Drosophila* reverses the default handedness in visceral organs. *Nature* 440, 798–802 (2006).
- Wood, W. B. Evidence from reversal of handedness in C. elegans embryos for early cell interactions determining cell fates. Nature 349, 536–538 (1991).
- Bergmann, D. C. et al. Embryonic handedness choice in C. elegans involves the Gα protein GPA-16. Development 130, 5731–5740 (2003).
- Boycott, A. E., Diver, C., Garstang, S. L., Hardy, A. C. & Turner, F. M. The inheritance of sinistrality in *Lymnaea peregra*. *Phil. Trans. R. Soc. Lond. B* 219, 51–131 (1930)
- Sturtevant, A. H. Inheritance of direction of coiling in *Lymnaea*. Science 58, 269–270 (1923).
- Freeman, G. & Lundelius, J. W. The developmental genetics of dextrality and sinistrality in the gastropod *Lymnaea peregra*. Wilhelm Roux Arch. Dev. Biol. 191, 69–83 (1982)
- Hosoiri, Y., Harada, Y. & Kuroda, R. Construction of a backcross progeny collection of dextral and sinistral individuals of a freshwater gastropod, *Lymnaea* stagnalis. Dev. Genes Evol. 213, 193–198 (2003).
- Crampton, H. E. Reversal of cleavage in a sinistral gastropod. Ann. NY Acad. Sci. 8, 167–170 (1894).
- Shibazaki, Y., Shimizu, M. & Kuroda, R. Body handedness is directed by genetically determined cytoskeletal dynamics in the early embryo. *Curr. Biol.* 14, 1462–1467 (2004).
- Meshcheryakov, V. N. & Beloussov, L. V. Asymmetrical rotations of blastomeres in early cleavage of gastropoda. Wilhelm Roux Arch. Dev. Biol. 177, 193–203 (1975).
- 22. Wandelt, J. & Nagy, L. M. Left-right asymmetry: more than one way to coil a shell. *Curr. Biol.* 14, R654–R656 (2004).

- Freeman, G. & Lundelius, J. W. Evolutionary implications of the mode of D quadrant specification in coelomates with spiral cleavage. *J. Evol. Biol.* 5, 205–247 (2002).
- 24. Duboc, V. & Lepage, T. A conserved role for the nodal signaling pathway in the establishment of dorso-ventral and left-right axes in deuterostomes. *J. Exp. Zool. B* 310, 41–53 (2008).
- Grande, C. & Patel, N. H. Nodal signalling is involved in left-right asymmetry in snails. *Nature* 457, 1007–1011 (2009).
- Nederbragt, A. J., van Loon, A. E. & Dictus, W. J. Expression of *Patella vulgata* orthologs of engrailed and *dpp-BMP2/4* in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. *Dev. Biol.* 246, 341–355 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Smit for his gift of dextral and sinistral stock of L. stagnalis. We also thank K. Miyoshi, Y. Ozawa and H. Kuwata of the Kuroda Chiromorphology team for their help in rearing snails and creating F_7 congenic snails. K. Fujikura, A. Okubo and G. Sai are thanked for their preliminary attempts at $in\ situ$ hybridization experiments.

Author Contributions R.K. conceived the study, designed/coordinated the experiments and wrote the manuscript. B.E. performed the reversal experiments and whole mount *in situ* hybridization (WISH) on reversed embryos. M.A. performed WISH on control and F_7 congenic snails. M.S. cloned and characterized *nodal* and *Pitx* from *L. stagnalis* to make template vector for the WISH probes. B.E. and M.S. provided comments on the manuscript.

Author Information Sequences of *L. stagnalis nodal* and *Pitx* are deposited at GenBank, with accession numbers respectively GU073383 and GU073384. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.K. (ckuroda@mail.ecc.u-tokyo.ac.jp).

doi:10.1038/nature08597 nature

METHODS

PCR amplification of Nodal sequence from *L. stagnalis* genomic DNA was accomplished using degenerate primers which were designed using the consensus-degenerate hybrid oligonucleotide (CODEHOP) program available at http://blocks.fhcrc.org/codehop.html. Degenerate PCR primers for amplifying Pitx sequence from *L. stagnalis* trochophore stage cDNA were designed as

described²⁷. To clone the full-length cDNA for the above genes, 5' and 3' RACE PCR was performed using the SMART-RACE kit from Clontech. Probes for *in situ* hybridization were designed using the Probe Search System available at http://probe-search.ccr.tokushima-u.ac.jp.

27. Christiaen, L. et al. Pitx genes in Tunicates provide new molecular insight into the evolutionary origin of pituitary. Gene 287, 107–113 (2002).

Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment

Robert L. Sinsabaugh¹, Brian H. Hill² & Jennifer J. Follstad Shah³

Biota can be described in terms of elemental composition, expressed as an atomic ratio of carbon:nitrogen:phosphorus (refs 1-3). The elemental stoichiometry of microoorganisms is fundamental for understanding the production dynamics and biogeochemical cycles of ecosystems because microbial biomass is the trophic base of detrital food webs⁴⁻⁶. Here we show that heterotrophic microbial communities of diverse composition from terrestrial soils and freshwater sediments share a common functional stoichiometry in relation to organic nutrient acquisition. The activities of four enzymes that catalyse the hydrolysis of assimilable products from the principal environmental sources of C, N and P show similar scaling relationships over several orders of magnitude, with a mean ratio for C:N:P activities near 1:1:1 in all habitats. We suggest that these ecoenzymatic ratios reflect the equilibria between the elemental composition of microbial biomass and detrital organic matter and the efficiencies of microbial nutrient assimilation and growth. Because ecoenzymatic activities intersect the stoichiometric and metabolic theories of ecology⁷⁻⁹, they provide a functional measure of the threshold at which control of community metabolism shifts from nutrient to energy flow.

Ecological stoichiometric theory uses elemental ratios and the concept of stoichiometric invariance to predict nutrient retention and biomass production on subcellular to ecosystem scales, creating a unifying theory of ecology⁵. The complementary metabolic theory of ecology uses thermodynamics and the concept of energetic invariance to predict metabolic activity across all levels of biological organization^{7–10}. Both theories have developed over the course of a century and are supported by large bodies of literature, but evaluations are complicated because ecological systems are controlled by non-equilibrium flows of both materials (nutrients) and energy (carbon) and it is difficult to know which is predominant in a specific context. Ecoenzyme (enzymes not contained by membranes of living cells) activity represents an intersection of ecological stoichiometric theory and metabolic theory of ecology because enzyme expression is a product of cellular metabolism specifically regulated by environmental nutrient availability (Supplementary Fig. 1).

Terrestrial soils and freshwater sediments contain reserves of organic carbon estimated at 1,500 petagrams (ref. 11; $1 \text{ Pg} = 10^{15} \text{ g}$) and 0.2 Pg (ref. 12), respectively. Mineralization of this organic matter by heterotrophic microorganisms creates the trophic base for detritus food webs, drives global carbon and nutrient cycles, and mediates plant

production and atmospheric composition. The proximate agents of organic matter decomposition are ecoenzymes that deconstruct plant and microbial cell walls, depolymerize macromolecules, and ultimately produce soluble substrates for microbial assimilation^{13,14}. Many of these enzymes are expressed and released into the environment by microorganisms in response to environmental signals; others enter the environment through cell lysis. The most widely assayed ecoenzymes are those involved in the degradation of lignocellulose, the principal product of primary production. Lignocellulose does not contain N or P so enzymes that hydrolyse proteins, chitin and peptidoglycan (the principal sources of organic N) and mineralize P from nucleic acids, phospholipids and other ester phosphates are also widely measured¹⁵.

The enzymatic degradation of biopolymers requires the synergistic interaction of several classes of enzymes. However, ecological studies generally quantify only the activities of enzymes that catalyse the terminal reactions that produce assimilable products from the principal C, N and P sources. The mostly widely measured activities include β -1,4-glucosidase (BG), β -1,4-N-acetylglucosaminidase (NAG), leucine aminopeptidase (LAP) and acid (alkaline) phosphatase (AP; Table 1). The potential activities of these enzymes are frequently linked to rates of microbial metabolism and biogeochemical processes, and used as indicators of microbial nutrient demand $^{15-17}$.

Microbial nutrient demand is determined by the elemental stoichiometry of microbial biomass in relation to environmental nutrient availability. The mean C:N:P ratio of plant litter is about 3,000:46:1 (ref. 2). Microbial respiration and nutrient immobilization increase the N and P content of litter as it decomposes, eventually creating soil organic matter with a mean C:N:P ratio of 186:13:1 (ref. 3). N and P are further concentrated in soil microbial biomass, which has a mean C:N:P ratio of 60:7:1 (ref. 3). Similar elemental ratios have been reported for heterotrophic microbial biomass associated with surface sediments of inland waters¹⁸.

We analysed the stoichiometry of BG, NAG, LAP and AP activities, using data collected in three synoptic studies of terrestrial soils and freshwater sediments (Table 2). These measures of organic C, N and P acquisition show similar scaling relationships that become increasingly correlated along the hydrologic gradient from terrestrial soil, to wetland (lentic) sediment, to river (lotic) sediment (Fig. 1). The slopes of the ecoenzymatic C:N and C:P regressions differ significantly by habitat, except for the C:N activity regressions for soil and lotic

Table 1 | Ecoenzymes included in this study

Enzyme	Abbrev.	EC*	Function
β-1,4-glucosidase β-1,4-N-acetylglucosaminidase Leucine aminopeptidase Acid (alkaline) phosphatase	BG NAG LAP AP	3.2.1.21 3.2.1.14 3.4.11.1 3.1.3.1	Cellulose degradation: hydrolyses glucose from cellobiose Chitin and peptidoglycan degradation: hydrolyses glucosamine from chitobiose Proteolysis: hydrolyses leucine and other hydrophobic amino acids from the N terminus of polypeptides Hydrolyses phosphate from phosphosaccarides and phospholipids

^{*} Enzyme Commission Number.

¹Biology Department, University of New Mexico, Albuquerque, New Mexico 871312, USA. ²US Environmental Protection Agency, National Health & Environmental Effects Laboratory, Duluth, Minnesota 55804-2595, USA. ³Biology Department, Duke University, Durham, North Carolina 27708, USA.

Table 2 | Mean ecoenzymatic activities of sediments and soils

	Habitat			
	Lotic	Lentic	Soil	
BG	7,580	20,090	4,960	
AP	5,080	39,500	7,980	
NAG	3,750	6,520	2,570	
LAP	1,580	46,400	950	
BG:AP	1.49	0.51	0.62	
BG:(NAG + LAP)	1.42	0.38	1.41	
(NAG + LAP):AP	1.05	1.34	0.44	
Sites	445	60	40	
Observations	445	60	935	

Lotic sediments were collected from the Upper Mississippi, Missouri and Ohio River systems using the US Environmental Protection Agency's EMAP synoptic sampling design; lentic sediments were collected from the Great Lakes Basin using the EMAP synoptic sampling design, and terrestrial soils were collected from 40 ecosystems that span all major biomes. Sediment and soil data from refs 17–19. Units of activity are nmol h $^{-1}$ per g C. BG:AP, BG:(NAG + LAP) and (NAG + LAP):AP are metrics for ecoenzymatic C:P, C:N and N:P acquisition activities, respectively. BG, β -1,4-glucosidase; NAG, β -1,4-N-acetylglucosaminidase; LAP, leucine aminopeptidase; AP, alkaline/acid phosphatase (AP).

sediment (Supplementary Information). Within habitats, C acquisition has similar scaling relationships with N and P acquisition (N/P 1.03/1.03 for lotic sediment, 0.77/0.76 for lentic sediment, 1.09/1.16 for soil, Fig. 1). Consequently, the slope of N:P activity regressions is approximately 1.0 for all habitats (Fig. 1). In all three environments, the mean ecoenzymatic C:N:P activity ratio is approximately 1:1:1 (Fig. 2). The mean C:P activity ratio for freshwater sediments is significantly greater than that for terrestrial soils (1.049 versus 0.959). The mean C:N activity ratio for terrestrial soils is significantly greater than that for freshwater sediments (1.023 versus 0.997). The stoichiometry among the C, N and P acquiring enzymes is more diffuse in terrestrial soils than in aquatic sediments (Fig. 1), probably because soils have greater heterogeneity in organic matter composition, nutrient distribution and edaphic conditions¹⁹ (Supplementary Information).

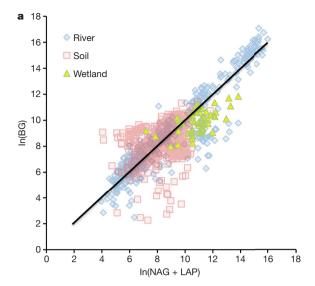
The distribution of ecoenzymatic C:N:P activity ratios identifies the boundaries of microbial community response to disturbances or fluctuations in nutrient availability, and shows that heterotrophic microbial communities of diverse composition share a common pattern of functional organization. The mean C:N:P stoichiometry of these enzyme activities indicates that the rates of supply of assimilable substrates from the respective C, N and P reservoirs are on average similar in magnitude, and thus that proximate limitations on microbial production may readily shift between C, N and P (ref. 20).

The similarity of ecoenzymatic C:N:P stoichiometry across ecosystems is unexpected, because these activities vary widely and experimental manipulations within ecosystems show that C, N and P acquisition activities can be modulated by inorganic nutrient availability^{21–23}. These responses are generally interpreted in the context of resource allocation models based on the premise that cellular resources directed towards N and P acquisition reduce resources available for C acquisition^{15,17} (Supplementary Information). Our analyses suggest that the capacity of microbial communities to alter relationships among ecoenzymatic C, N and P acquisition activities in response to environmental resource availability is limited, with similar boundaries in all habitats.

The stoichiometry of ecoenzymatic activity can be related to ecological stoichiometric theory and the metabolic theory of ecology through the threshold elemental ratio (TER) concept^{9,24}. TER is the critical elemental C:P or C:N ratio at which metabolic control of an ecological system switches from energy flow, represented by C, to limiting nutrient flow, represented by P or N. TER for P and N can be represented as⁹:

$$TER_{C:P} = \left(\frac{A_{P}}{GE}\right) B_{C:P}$$
 (1)

$$TER_{C:N} = \left(\frac{A_{N}}{GE}\right) B_{C:N} \tag{2}$$



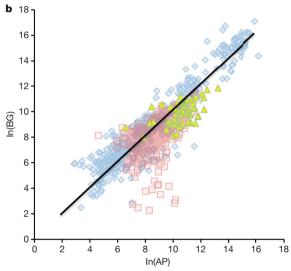


Figure 1 Organic nitrogen (N) acquisition activity and organic phosphorus (P) acquisition activity in relation to carbon (C) acquisition. N acquisition is measured by the potential activities of β-1,4-N-acetylglucosaminidase (NAG) and leucine aminopeptidase (LAP) (a); P acquisition is measured as acid (alkaline) phosphatase (AP) activity (b); C acquisition is represented by β-1,4-glucosidase (BG). For lotic sediments, the standardized major axis (type II) regression for C:P acquisition activity is $ln(BG) = 1.030 \times ln(AP) + 0.149, R^2 = 0.897, P < 0.001$; the C:N regression is $ln(BG) = 1.032 \times ln(NAG + LAP) - 0.268$, $R^2 = 0.882$, P < 0.001; n = 445. For lentic sediments, the type II C:P regression is $ln(BG) = 0.763 \times ln(AP) + 1.831, R^2 = 0.330, P < 0.001$; the C:N regression is $\ln(BG) = 0.767 \times \ln(NAG + LAP) + 1.520$, $R^2 = 0.323$, P < 0.001; n = 60. For soils the C:P regression is $ln(BG) = 1.162 \times ln(AP) - 1.845$, $R^2 = 0.396$, P < 0.001; the C:N regression is $ln(BG) = 1.091 \times ln(NAG + LAP) - 0.660$, $R^2 = 0.157$, P < 0.001; n = 929). The slopes of enzymatic C:P and C:N regressions differ significantly by habitat (P < 0.017) with the exception of C:N regression slopes for soil and lotic sediments. For the global data set, the type II C:P regression is $ln(BG) = 1.037 \times ln(AP) - 0.481$, $R^2 = 0.761$, P < 0.001; the C:N regression is $ln(BG) = 1.018 \times ln(NAG + LAP) - 0.124$, $R^2 = 0.695$, P < 0.001; n = 1434); the N:P regression is $\ln(\text{NAG} + \text{LAP}) = 1.018 \times \ln(\text{AP}) - 0.351, R^2 = 0.694, P < 0.001.$ Reference lines with a slope of 1.0 are shown on the graphs.

where $A_{\rm P}$ and $A_{\rm N}$ are assimilation efficiencies for P and N, GE is microbial growth efficiency and $B_{\rm C:P}$ and $B_{\rm C:N}$ are the C:P and C:N ratios of microbial biomass. Using the mean elemental C:P ratios for soil organic matter (186) and microbial biomass (60; ref. 3) as values for TER_{C:P} and $B_{\rm C:P}$, and 0.9 for $A_{\rm P}$ (ref. 9), mean microbial growth efficiency is 0.29, the same value reported for aquatic invertebrates

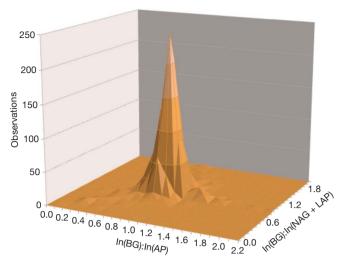


Figure 2 | **Frequency distribution of ecoenzymatic ratios.** Distribution of heterotrophic microbial communities in terrestrial soils and freshwater sediments in relation to ecoenzymatic C:N and C:P acquisition activity, as measured by ratios of BG:(NAG + LAP) and BG:AP activities. By these measures, the mean global ratio of C:N:P acquisition is approximately 1:1:1 (mean C:N 1.01 (s.d. 0.15), mean C:P 0.99 (0.14)). Mean values for lotic sediment, wetland sediment and soil C:P activity are 1.063 (0.170), 0.943 (0.098) and 0.959 (0.114), respectively; mean values for C:N acquisition activity are 1.008 (0.130), 0.912 (0.091) and 1.023 (0.163), respectively. The mean ecoenzymatic C:P acquisition ratio for freshwater sediment (1.049) is significantly greater than that for soils (analysis of variance, $F_{1,1440} = 147$, $P < 10^{-31}$). The mean ecoenzymatic C:N acquisition ratio for freshwater sediment (0.997) is significantly lower than that of soils ($F_{1,1440} = 10.6$, P = 0.011).

and similar to mean growth efficiencies for soil²⁵ and freshwater²⁶ microorganisms. The calculation suggests that C:P ratios >186 and GE < 0.29 are indicators of P limitation, rather than energy (C) limitation. The parallel calculation using elemental C:N ratios of 14.3 and 8.6 for soil organic matter and microbial biomass³ yields a GE value of 0.54. Because this value does not correspond with the GE estimate from the TER_{C:P} calculation and lies near the extreme for microbial community GE estimates, we infer either that mean A_N is half that of A_P , or more likely, that the mean TER_{C:N} is greater than the mean C:N ratio of soil organic matter. If $A_N = A_P = 0.9$ and mean GE is 0.29, then TER_{C:N} is approximately 27. This value is consistent with many empirical estimates of TER_{C:N} (refs 16, 27) and with analyses indicating that TER values are 2–3 times greater than elemental biomass ratios²⁴.

Ratios of ecoenzymatic activities can be related to both resource availability and microbial metabolism:

$$\text{TER}_{\text{C:P}}/L_{\text{C:P}} \propto \text{BG/AP} \propto L_{\text{C:P}}/B_{\text{C:P}}$$
 (3)

$$\text{TER}_{\text{C:N}}/L_{\text{C:N}} \propto \text{BG/(NAG} + \text{LAP}) \propto L_{\text{C:N}}/B_{\text{C:N}}$$
 (4)

where $L_{\text{C:P}}$ and $L_{\text{C:N}}$ are the elemental C:P and C:N ratios of labile organic matter, that is, organic matter readily available for microbial consumption. Equations (3) and (4) can be reduced and combined with equations (1) and (2):

$$BG/AP = p_0(TER_{C:P}/B_{C:P}) = p_0(A_P/GE)$$
 (5)

$$BG/(NAG + LAP) = n_0(TER_{C:N}/B_{C:N}) = n_0(A_N/GE)$$
 (6)

where p_0 and n_0 are normalization constants (Supplementary Information).

Using the mean C:P value of soil organic matter (186) for TER_{C:P}, the mean C:P of microbial biomass (60) for $B_{\text{C:P}}$, and a mean BG:AP ratio of 1.0, the log_e transformed version of equation (5) returns a $\ln(p_0)$ of

-1.13. Alternatively, the observed $\ln(p_0)$ of -1.84 from the soil C:P regression (Fig. 1) returns a BG:AP ratio of 0.5 (Table 1, equivalent to $\ln(\text{BG})/\ln(\text{AP})$ of 0.95). For nitrogen, using a TER_{C:N} estimate of 27, the mean C:N of microbial biomass (8.6) for $B_{\text{C:N}}$ and a mean BG:(NAG + LAP) ratio of 1.0, the \log_{e} transformed version of equation (6) returns a $\ln(n_0)$ of -1.14; using the observed $\ln(n_0)$ of -0.66 from the soil C:N regression (Fig. 1) returns a BG/(NAG + LAP) ratio of 1.6 (Table 2, equivalent to $\ln(\text{BG})/\ln(\text{NAG} + \text{LAP})$ of 1.06).

At a global scale, these relationships show that ecoenzymatic activity ratios can be empirically linked to the biogeochemical equilibrium between microbial growth efficiency and the elemental compositions of soil organic matter and microbial biomass. Because ecoenzymatic activities are related to both resource availability and microbial growth on the same spatiotemporal scale, analyses of these activities may be used to resolve energetic and nutrient constraints on microbial community metabolism in the context of stoichiometric and metabolic theories of ecology.

METHODS SUMMARY

The soil data were obtained from ref. 19; the freshwater sediment data were presented in refs 28 and 29 (Table 2). A common assay protocol using fluorigenic substrates was used in all studies. Cases with missing values for one or more enzyme activity were excluded from our analyses. Where necessary, activities expressed per g organic matter (OM) were converted to units per g C using a conversion of 0.45 g C/g OM. The type II regression results (Fig. 1) were calculated using SMATR v2.0³⁰ (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 July; accepted 29 October 2009.

- Redfield, A. The biological control of chemical factors in the environment. Am. Sci. 46, 205–221 (1958).
- Reich, P. B. & Oleksyn, J. Global patterns of plant leaf N and P in relation to temperature and latitude. Proc. Natl Acad. Sci. USA 101, 11001–11006 (2004).
- Cleveland, C. C. & Liptzin, D. C:N:P stoichiometry in soil: is there a "Redfield ratio" for the microbial biomass? *Biogeochemistry* 85, 235–252 (2006).
- Elser, J. J. et al. Nutritional constraints in terrestrial and freshwater foodwebs. Nature 408, 578–580 (2000).
- Sterner, R. W. & Elser, J. J. Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere (Princeton Univ. Press, 2002).
- Manzoni, S., Jackson, R. B., Trofymow, J. A. & Porporato, A. The global stoichiometry of litter nitrogen mineralization. *Science* 321, 684–686 (2008).
- Gillooly, J. F. et al. The metabolic basis of whole-organism RNA and phosphorus content. *Proc. Natl Acad. Sci. USA* 102, 11923–11927 (2005).
- 8. Kerkoff, A. J., Enquist, B. J., Elser, J. J. & Fagan, W. F. Plant allometry, stoichiometry and temperature-dependence of primary productivity. *Glob. Ecol. Biogeogr.* 14, 585–598 (2005).
- Allen, A. P. & Gillooly, J. F. Towards an integration of ecological stoichiometry and the metabolic theory of ecology to better understand nutrient cycling. *Ecol. Lett.* 12, 369–384 (2009).
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* 85, 1771–1789 (2004).
- Houghton, R. A. Balancing the global carbon budget. Annu. Rev. Earth Planet. Sci. 35, 313–347 (2007).
- 12. Cole, J. et al. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* 10, 171–184 (2007).
- 13. Chróst, R. J. Microbial Enzymes in Aquatic Environments (Springer, 1991).
- 14. Burns, R. G. & Dick, R. P. Enzymes in the Environment: Activity, Ecology and Applications (Dekker, 2002).
- Allison, S. D., Gartner, T., Holland, K., Weintraub, M. & Sinsabaugh, R. L. in Manual of Environmental Microbiology (eds Hurst, C. J., Knudsen, G. R., McInerney, M. J., Stetzenbach, L. D. & Walter, M. V.) 704–711 (ASM Press, 2007).
- Schimel, J. P. & Weintraub, M. N. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. Soil Biol. Biochem. 35, 549–563 (2003).
- Moorhead, D. L. & Sinsabaugh, R. L. A theoretical model of litter decay and microbial interaction. *Ecol. Monogr.* 76, 151–174 (2006).
- Cross, W. F., Benstead, J. P., Frost, P. C. & Thomas, S. A. Ecological stoichiometry in freshwater benthic systems: recent progress and perspectives. *Freshwat. Biol.* 50, 1895–1912 (2005).
- Sinsabaugh, R. L. et al. Stoichiometry of soil enzyme activity at global scale. Ecol. Lett. 11, 1252–1264 (2008).
- Sinsabaugh, R. L. & Follstad Shah, J. J. Integrating resource utilization and temperature in metabolic scaling of riverine bacterial production. *Ecol. Monogr.* (in the press).

- 21. Olander, L. P. & Vitousek, P. M. Regulation of soil phosphatase and chitinase activity by N and P availability. *Biogeochemistry* **49**, 175–190 (2000).
- Sinsabaugh, R. L., Carreiro, M. M. & Repert, D. A. Allocation of extracellular enzymatic activity in relation to litter composition, N deposition, and mass loss. *Biogeochemistry* 60, 1–24 (2002).
- Sinsabaugh, R. L., Gallo, M. E., Lauber, C., Waldrop, M. & Zak, D. R. Extracellular enzyme activities and soil carbon dynamics for northern hardwood forests receiving simulated nitrogen deposition. *Biogeochemistry* 75, 201–215 (2005).
- 24. Frost, P. C. et al. Threshold elemental ratios of carbon and phosphorus in aquatic consumers. Ecol. Lett. 9, 774–779 (2006).
- Herron, P. M., Stark, J. M., Holt, C., Hooker, T. & Cardon, Z. G. Microbial growth efficiencies across a soil moisture gradient assessed using ¹³C-acetic acid vapor and ¹⁵N-ammonia gas. Soil Biol. Biochem. 41, 1262–1269 (2009).
- Sand-Jensen, K., Pedersen, N. L. & Sondergaard, M. Bacterial metabolism in small temperate streams under contemporary and future climates. Freshwat. Biol. 52, 2340–2353 (2007).
- 27. Parton, W. et al. Global-scale similarities in nitrogen release patterns during long term decomposition. *Science* **315**, 361–362 (2007).
- 28. Hill, B. H. et al. Sediment microbial enzyme activity as an indicator of nutrient limitation in Great Lakes coastal wetlands. Freshwat. Biol. 51, 1670–1683 (2006).

- Hill, B. H., Elonen, C. M., Jicha, T. M., Bolgrien, D. W. & Moffett, M. F. Sediment microbial enzyme activity as an indicator of nutrient limitation in the great rivers of the Upper Mississippi River basin. *Biogeochemistry*. doi:10.1007/s10533-009-9366-0 (2009).
- 30. Warton, D. I., Wright, J., Falster, D. S. & Westoby, M. Bivariate line-fitting methods for allometry. *Biol. Rev. Camb. Phil. Soc.* **81,** 259–291 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.J.F.S. was supported by the National Science Foundation (DBI-0630558).

Author Contributions R.L.S. provided data from soils. B.H.H. provided data from freshwater sediments. J.J.F.S. and R.L.S. collaborated on data synthesis and interpretation. All authors contributed to manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.L.S. (rlsinsab@unm.edu).

doi:10.1038/nature08632 nature

METHODS

Sample collection. Terrestrial soil, excluding surface litter, was collected at each site to depths of 5–20 cm. The samples included in this study are a subsample of those assembled in ref. 19 for a meta-analysis of soil enzyme activities and include all major biomes. Nearly all of the sampling was done in the past ten years. Sampling design varied among ecosystems. Metadata for the individual terrestrial ecosystems included in Table 2 were presented in ref. 19.

River sediment samples were collected as part of the US Environmental Protection Agency's Environmental Monitoring and Assessment Program for Great Rivers (EMAP-GRE), which includes the Upper Mississippi River from Lower St Anthony Falls in Minneapolis-St Paul, Minnesota, to the confluence with the Ohio River at Cairo, Illinois; the Missouri River from Fort Peck Dam in Montana to the confluence with the Mississippi River at St Louis, Missouri; and the Ohio River from the confluence of the Allegheny and Monongahela Rivers in Pittsburgh, Pennsylvania, to the confluence with the Mississippi River The EMAP-GRE design was spatially balanced and used an unequal probability for selecting sites based on unique river reaches, for example, the unimpounded Upper Mississippi River. The survey design selected 447 sites representing 4,838 km of river channel. Sampling at these sites occurred July–September 2004–06. Surface sediment samples (top 5 cm) at each site were collected at 11

stations equally spaced along a 500 m, longitudinal channel-margin transect and combined for all stations at a site, resulting in a single composite sample per site.

Lentic sediment samples were collected from coastal wetlands within the Laurentian Great Lakes²⁸. The sampled wetlands represent the basin-wide range of natural conditions and human impacts on wetlands. Twenty-four wetlands were sampled in 2002, 20 in 2003 and 14 in 2004. Sampling locations within each wetland were selected by dividing the wetted perimeter of the wetland into seven equal interval segments with sampling stations corresponding to segment boundaries. Surface sediment samples (top 5 cm) were collected from emergent vegetation, submerged vegetation and open water zones. Within each zone sediment collections were combined for all stations resulting in three samples per wetland. Cases were included in our analyses only if they had values for all four ecoenzymatic activities (Table 2).

Enzyme assays. Hydrolytic enzyme activities were assayed using substrates linked to a methylumbelliferyl fluor^{19,28,29}. Sediment samples were assayed at 20 °C using pH 8 bicarbonate buffer. Soil samples were assayed at 20 °C in either pH 8 bicarbonate buffer or pH 5 acetate buffer to approximate the bulk soil pH of the system from which they were collected. Activities were calculated in units of nmol h⁻¹ per g C. **Statistical analyses.** Statistical analyses and results are described in the text with additional results included in the Supplementary Information.

LETTERS

Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target

Szymon Swiezewski¹, Fuquan Liu¹, Andreas Magusin¹ & Caroline Dean¹

Transcription in eukaryotic genomes generates an extensive array of non-protein-coding RNA, the functional significance of which is mostly unknown¹. We are investigating the link between noncoding RNA and chromatin regulation through analysis of FLC a regulator of flowering time in Arabidopsis and a target of several chromatin pathways. Here we use an unbiased strategy to characterize non-coding transcripts of FLC and show that sense/antisense transcript levels correlate in a range of mutants and treatments, but change independently in cold-treated plants. Prolonged cold epigenetically silences FLC in a Polycombmediated process called vernalization2. Our data indicate that upregulation of long non-coding antisense transcripts covering the entire FLC locus may be part of the cold-sensing mechanism. Induction of these antisense transcripts occurs earlier than, and is independent of, other vernalization markers3 and coincides with a reduction in sense transcription. We show that addition of the FLC antisense promoter sequences to a reporter gene is sufficient to confer cold-induced silencing of the reporter. Our data indicate that cold-induced FLC antisense transcripts have an early role in the epigenetic silencing of FLC, acting to silence FLC transcription transiently. Recruitment of the Polycomb machinery then confers the epigenetic memory. Antisense transcription events originating from 3' ends of genes might be a general mechanism to regulate the corresponding sense transcription in a condition/ stage-dependent manner.

Advances in sequencing technology have enabled the discovery of many long non-coding transcripts. Their functions and evolution are poorly understood¹, but many seem to be antisense transcripts that initiate and terminate near the terminators and promoters, respectively, of sense transcripts⁴. We have previously found non-coding transcripts originating from the 3' region of the gene encoding the floral repressor FLOWERING LOCUS C (FLC)^{5,6}. FLC is upregulated by FRIGIDA and transcriptionally silenced through the activity of an RNA recognition motif (RRM) protein (FCA) and the K4 histone demethylase activity of FLOWERING LOCUS D (FLD)5,6. FLC is also repressed by prolonged cold and epigenetically silenced by a Polycomb mechanism involving a PHD-PRC2 (plant homeodomain-Polycomb repressive complex 2) complex2, in a process known as vernalization^{3,7}. To understand the roles of non-coding RNA in these different types of FLC regulation we undertook an unbiased analysis of FLC transcripts in a selection of genotypes and environments. We designed a single nucleotide resolution array (NimbleGen/Affymetrix platform) of 25-base oligonucleotides covering both strands of FLC and the surrounding 50 kb region. We hybridized the arrays with total RNA (ribosomal RNA reduced through the use of Invitrogen RiboMinus—see Supplementary Methods) isolated from the following genotypes and treatments; Columbia (wild type), FRIGIDA+ (FRI), fca, 35S::FCAy (an overexpression of FCA), FRI seedlings that had been cold-treated for

14 days, FRI seedlings that had been cold-treated then followed by 7 days of further growth at 20 °C. Fourteen days of cold was chosen with the aim of identifying the early steps in vernalization; this is insufficient to saturate the vernalization requirement but does result in stable reduction in FLC expression.

Despite considerable differences in hybridization efficiencies between the different oligonucleotides there was excellent reproducibility between biological replicates (Fig. 1a, b). Comparison of the hybridization profile between the different genotypes and environments confirmed all previous findings of FRIGIDA and FCA function on FLC expression^{8,9}, so validating the approach. Means of normalized hybridization data for each oligonucleotide in each treatment are shown in Figure 1a, b (FLC) and Supplementary Fig. 1 (whole 50 kb). Extensive FLC antisense transcripts were detected in all genotypes and treatments with levels correlating positively with sense transcript levels across all but one of the comparisons (Fig. 1). Additional quantitative PCR with reverse transcription (qRT–PCR) analysis of sense/antisense levels corroborated the array data and analysis of additional autonomous pathway mutants confirmed the close connection between sense and antisense levels (Supplementary Fig. 2e). This is consistent with the observation from many genomewide studies showing a positive correlation in expression of most sense/antisense pairs 10,11. The exception was cold-treated seedlings in which the ratio of sense to antisense FLC transcript changed significantly; antisense transcript accumulated over the majority of FLC, extending beyond the FLC 5' start and 3' polyadenylation sites, whereas levels of sense mRNA did not change significantly (Fig. 2a). Cold-induced accumulation of antisense transcripts for the other ten genes on the array was not observed (Supplementary Fig. 1). The accumulation of antisense FLC RNA was transient; levels returned to almost non-vernalized levels in plants returned to the warm for 7 days following 14 days cold (Fig. 1c). To determine the structure of the antisense transcripts we performed RACE (rapid amplification of cDNA ends) experiments. These defined several, cold-induced antisense FLC transcripts, with heterogeneous 5' ends initiating in a 100-nucleotide region immediately 3' to the sense FLC polyadenylation site (Fig. 2b, Supplementary Fig. 2). This coincides with the localized small RNA-induced heterochromatic region at the 3' end of FLC we have previously described⁵. As a parallel to HOTAIR (Hox antisense intergenic RNA)¹² we have called FLC antisense transcripts initiating from this region COOLAIR (cold induced long antisense intragenic RNA). RACE experiments showed that a proportion of the transcripts are either capped or carry a 5' triphosphate group, have long (60-100 adenosines) poly-A tails and are spliced (Supplementary Fig. 2b, c). We find alternatively processed forms, with respect to both polyadenylation site and splicing. Our data indicate that these antisense transcripts are produced from a DNA template as the splice sites are canonical and do not overlap with sense transcript splice sites. The antisense FLC transcripts also

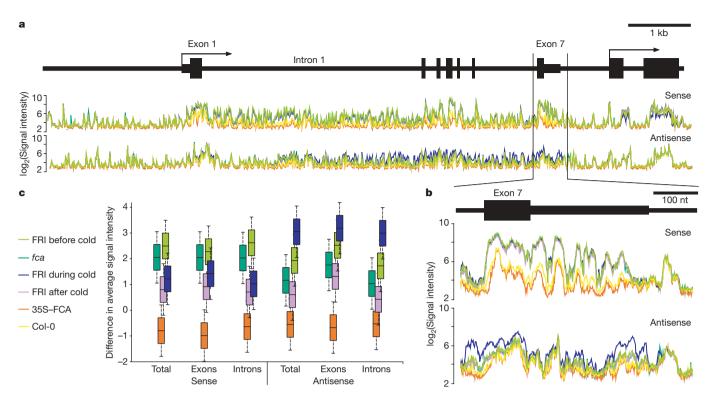


Figure 1 | *FLC* is associated with extensive antisense transcription. **a**, Signal intensity (log2) for the median of a 10-base-pairs staggered window along the genomic region of *Arabidopsis* chromosome V containing *FLC*. Genomic structure of *FLC*; black boxes represent exons; lines represent IGR and introns. **b**, Magnified region of *FLC* 3' end showing *FLC* antisense transcript start. **c**, Boxplots illustrating the difference in average signal

intensity for sense and antisense strands compared to Col-0 (Columbia wild type). The central box represents the data between the quartiles, the median is shown by the line through the centre of box; and the whiskers extend out to the extremes of the data. Total, whole *FLC* gene; exons, sense *FLC* exons, introns, sense *FLC* introns.

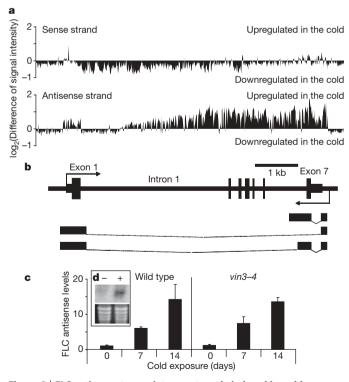


Figure 2 | *FLC* antisense transcripts are strongly induced by cold. **a**, Difference of signal intensities (log2) for cold-treated and untreated plants along sense and antisense strands of *FLC*. **b**, *FLC* region as in Fig. 1a with antisense transcript structure shown below. **c**, qPCR quantification of relative *FLC* antisense transcript levels in wild type and vin3-4. Data are means and s.e.m. (n = 2). **d**, Northern blot of of *FLC* antisense transcripts in control seedlings (-) and those cold-treated for 14 days (+).

extend beyond each end of the sense transcript (Fig. 2b, Supplementary Fig. 2a,d).

Given that Polycomb silencing of FLC is cold-dependent we explored the timing of COOLAIR induction relative to other known markers in the vernalization process. The earliest reported step to date is the cold-induction of VIN3 (VERNALIZATION INSENSITIVE 3), a PHD protein required for vernalization³. VIN3 expression is first detectable 20 days after transfer to cold, is maximal 40 days after transfer and is then undetectable 3 days after return to warm conditions³. VIN3 heterodimerizes with a related PHD protein VRN5 (VERNALIZATION 5, ref. 13) and induces formation of a PHD-PRC2 complex that associates with a specific domain in the first intron of FLC after prolonged cold2. This association triggers spreading of VRN5 along FLC and significant enhancement of histone H3 Lys 27 trimethylation (H3K27me3) levels throughout the locus, to levels required for maintenance of the silencing during subsequent growth of the plant². COOLAIR induction was independent of VIN3 (Fig. 2c) and VIN3 induction was found to be much later than maximal COOLAIR induction (Fig. 4c). Because COOLAIR induction might precede the nucleation of the Polycomb silencing at FLC, we analysed the timing of repression of FLC transcription by cold. This was achieved in three ways; analysis of the exon/intron hybridization to the array, through quantitative PCR analysis of nascent and spliced sense FLC transcripts, and using chromatin immunoprecipitation experiments with antibodies that react with all forms of RNA polymerase II (RNA polII) (Fig. 1c, 3a, 3b). Using all three assays FLC sense transcription was found to decrease much faster than spliced sense transcript after the plants were transferred to cold (Fig. 3b), and this was again independent of VIN3 (Fig. 3c). Thus, an early step in vernalization seems to be cold-induced upregulation of antisense transcription linked to sense transcriptional repression that is independent of, and precedes, the maximum induction of VIN3. This is reminiscent of bxd silencing where non-coding RNA transcription has been shown to be involved in its 'early' repression

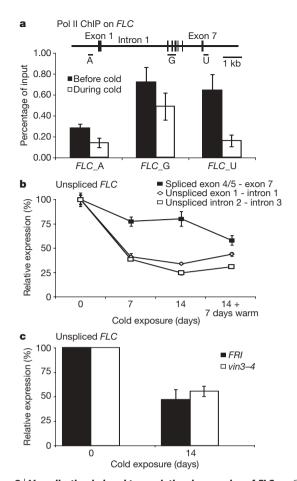


Figure 3 | **Vernalization-induced transcriptional repression of FLC. a**, ChIP using RNA polII-specific antibody. Regions of *FLC* monitored in the ChIP are shown as bars below a schematic of the genomic structure of *FLC*. Controls for the ChIP are in Supplementary Information. Data are means and s.e.m. (n=3). **b**, qPCR quantification of *FLC* mRNA and *FLC* unspliced sense transcript after different cold treatments. Data are means and s.e.m. (n=2). **c**, qPCR quantification of *FLC* unspliced sense transcript showing cold-induced downregulation is unaffected in *vin3*–4. Data are means and s.e.m. (n=2).

before Polycomb action¹⁴. The slower reduction of *FLC* spliced sense transcript may indicate a stable pool of reserve transcript maintained after transfer to cold that decreases after several weeks of cold, coincident with the decrease in antisense transcription (Fig. 4c).

FLC sense transcriptional silencing coincided with maximal COOLAIR expression so we tested whether COOLAIR was sufficient to cause transcriptional silencing. The COOLAIR promoter (aka FLC sense 3' region) was fused downstream of the coding region of a 35S (strong constitutive) promoter-green fluorescent protein (GFP) fusion (Fig. 4a). The 3' region of a gene showing no antisense transcription (rbcS3A) was used as control. There was no consistent change in expression of the sensor carrying the rbcS3A 3' region; whereas the COOLAIR promoter was found to reproducibly confer cold-dependent silencing of the sensor construct (a representative set from a larger number of transformants is shown in Fig. 4a). This was associated with cold-induction of an antisense transcript to the GFP sensor construct, showed by strand-specific RT-PCR analysis (Fig. 4b). Sequencing of the PCR products showed the transcript to be spliced at the same position as COOLAIR (Supplementary Fig. 2d) and extend from the COOLAIR promoter to at least half way through the transgene (Fig. 4a). The cold-induced silencing of the GFP sensor construct was transient and expression increased once plants were returned to warm conditions (Supplementary Fig. 3). This reinforces the view that it is the process of transcription rather than specific sequences within the FLC antisense transcripts that cause the cold-induced silencing. Supporting this view, we have found the

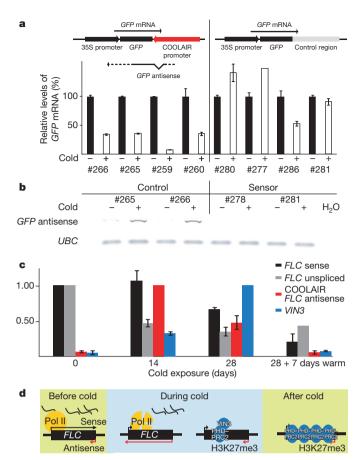


Figure 4 | FLC antisense promoter region is sufficient to confer cold-dependent silencing of sense transcription. a, Schematic representation and expression of GFP sensor lines. Warm-grown (black) or 14 days cold-treated (white). Structure of GFP antisense transcript shown below. b, GFP antisense transcript levels in GFP sensor lines. GFP sense expression data for line #278 shown in Supplementary Fig. 4. UBC, Arabidopsis ubiquitin-conjugating enzyme gene. c, Relative temporal expression of spliced and unspliced FLC sense transcript, COOLAIR and VIN3 before cold, after 14 days and 28 days of cold and then 7 days after transfer back to warm. Data are means and s.e.m. (n=2). d, Model indicating sequence of events. Green/blue, warm/cold, respectively.

COOLAIR promoter alone can drive cold-induced expression of LUC sequences (Supplementary Fig. 4). In the context of the endogenous *FLC* gene, the cold-induced increase in the antisense transcripts is linked to reduction in levels of nascent sense transcript and reduction in occupancy of RNA polII at *FLC*. This, together with the fact that COOLAIR transcripts extend beyond the transcription start site of *FLC*, indicates that antisense transcription might contribute to *FLC* transcriptional repression through promoter interference (Fig. 4d).

A similar cold-induced transcriptional silencing has been described for PHO84 in S. cerevisiae15. Cold-induced exosomedependent stabilization of antisense transcripts led to histone deacetylase recruitment and stable PHO84 silencing. Although we cannot exclude a cold-dependent post-transcriptional stabilization of COOLAIR, specific sequences or RNA structures seem to be dispensable as the promoter region alone is sufficient for cold-induced transcript production. However, we also see a consistent cold-dependent decrease in RNA polII occupancy at the COOLAIR promoter and elsewhere in FLC. This could indicate involvement of post-transcriptional mechanisms in COOLAIR induction or involvement of an alternative DNA-dependent RNA polymerase (although induction still occurs in a polIV/polV double mutant¹⁶; Supplementary Fig. 5). It may also simply reflect that the reduction in sense transcription is masking any increase in RNA polII occupancy associated with increased antisense transcription.

In summary, our study has demonstrated the environmental induction of long, non-coding, antisense transcripts covering an entire gene. These were found to be sufficient to induce silencing of a linked reporter. In the context of endogenous FLC silencing we propose that COOLAIR is involved in early, cold-dependent and transient, transcriptional silencing of FLC. This silencing would then be reinforced and epigenetically maintained by the Polycomb machinery. A possible sequence of events during the cold might be COOLAIR induction, suppression of sense transcription, VIN3 upregulation and PHD-PRC2 complex nucleation². In agreement with this we find COOLAIR expression continues longer in the cold in the absence of VIN3, supporting a model where the Polycomb machinery acts after COOLAIR and silences the whole FLC locus preventing further sense and antisense transcription (Supplementary Fig. 6). Recent data on PRC2 complex interaction with non-coding RNA transcripts12,17 might indicate that COOLAIR has additional, Polycomb-dependent functions in vernalization. Cold-responsiveness of an FLC-GUS transgene lacking the COOLAIR sequences and a possible Polycomb-independent silencing pathway that might operate in trans¹⁸ may also indicate mechanistic redundancy between different vernalization pathways. Our data are consistent with COOLAIR transcription being sufficient to silence linked sequences in a Polycomb-independent manner; however, these non-coding RNAs may function in both *cis* and *trans*¹⁹.

In a wider context, our discoveries on *FLC* may have broad relevance to other organisms where a genome-wide association of non-coding RNA transcript start sites with 3' regions has been observed⁴. Many studies have found a strong correlation between sense and antisense transcript abundance^{10,11}. On the basis of our data we speculate that many antisense transcription events originating from the 3' end of genes might regulate the sense partner in a condition/stage dependent manner. Continuous production of low levels of non-coding antisense transcripts maybe the cost of this regulatory mechanism.

METHODS SUMMARY

Plant growth conditions and vernalization treatment. All genotypes used have been described previously 8,16,20 . All seeds were surface-sterilized and grown on MS medium minus glucose. The 14-day cold-treatment was carried out on plants pre-grown for 7 days at 20 °C. Cold treatment was at $\sim\!5$ °C under short-day conditions, 8 h white light (10 mmol photons $\rm m^{-2}\,s^{-1}$), 16 h dark. Seedlings were then either harvested directly or transferred to long-day conditions of 16 h white light (57 mmol photons $\rm m^{-2}\,s^{-1}$), 8 h dark and grown for 7 days at 20 °C. Non-vernalized seedlings were cold-treated for 2 days, then transferred to long-day conditions for 14 days.

Array. Total RNA that was rRNA-depleted was labelled and hybridized by Cogenics on custom-design Affymetrix/Nimble Express GeneChips. For description of array design and analysis see Methods. COOLAIR sequences have been deposited in GenBank under accession numbers GQ352646 and GQ342259. Array data have been deposited to the Gene Expression Omnibus repository under the GEO ID GSE16977.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 July; accepted 30 October 2009.

 Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. Cell 136, 629–641 (2009).

- De Lucia, F. et al. A PHD-Polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc. Natl Acad. Sci. USA 105, 16831–16836 (2008).
- Sung, S. & Amasino, R. M. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. Nature 427, 159–164 (2004).
- 4. He, Y. et al. The antisense transcriptomes of human cells. Science 322, 1855–1857 (2008).
- Swiezewski, S. et al. Small RNA-mediated chromatin silencing directed to the 3' region of the Arabidopsis gene encoding the developmental regulator, FLC. Proc. Natl Acad. Sci. USA 104, 3633–3638 (2007).
- Liu, F. et al. The Arabidopsis RNA-binding protein FCA requires a lysinespecific demethylase 1 homolog to downregulate FLC. Mol. Cell 28, 398–407 (2007)
- Bastow, R. et al. Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427, 164–167 (2004).
- 8. Lee, I., Michaels, S. D., Masshardt, A. S. & Amasino, R. M. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of *Arabidopsis*. *Plant J.* **6**, 903–909 (1994).
- 9. Macknight, R. et al. FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell 89, 737–745 (1997).
- Oeder, S., Mages, J., Flicek, P. & Lang, R. Uncovering information on expression of natural antisense transcripts in Affymetrix MOE430 datasets. BMC Genomics 8, 200 (2007)
- RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and the FANTOM Consortium. Antisense transcription in the mammalian transcriptome. Science 309, 1564–1566 (2005).
- 12. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
- 13. Greb, T. et al. The PHD finger protein VRN5 functions in the epigenetic silencing of *Arabidopsis FLC. Curr. Biol.* 17, 73–78 (2007).
- 14. Petruk, S. et al. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell 127, 1209–1221 (2006).
- Camblong, J. et al. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell 131, 706–717 (2007).
- Onodera, Y. et al. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120, 613–622 (2005).
- Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell 32, 232–246 (2008)
- Sheldon, C. C., Finnegan, J. E. & Peacock, W. J. and Dennis, E. S., Mechanisms of gene repression by vernalization in *Arabidopsis. Plant J.* 59, 488–498 (2009).
- Camblong, J. et al. Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. Genes Dev. 23, 1534–1545 (2009).
- Sung, S., Schmitz, R. J. & Amasino, R. M. A PHD finger protein involved in both the vernalization and photoperiod pathways in *Arabidopsis*. *Genes Dev.* 20, 3244–3248 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all members of the Dean laboratory for discussions, J. Ecker and B. Gregory (Salk Institute) for help with the RiboMinus technique, C. Lister for help with cloning steps, M. Lenhard and H. Breuninger for seeds of LUC control plants. This research was supported by the BBSRC Core Strategic Grant to the John Innes Centre and BBSRC BB/D010799/1, NERC NE/C507629 and EU SIROCCO LSHG-CT-2006-037900 grants to C.D.

Author Contributions C.D. and S.S. designed the study. S.S. performed most experiments. S.S. and F.L. performed the array experiment. A.M. and S.S. analysed microarray data. C.D. and S.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.D. (caroline.dean@bbsrc.ac.uk).

doi:10.1038/nature08618 nature

METHODS

RNA quantification. For the *FLC* antisense northern blot a probe was generated using PCR on genomic DNA with the primers FLCexon7 5'-GGAGAATAATCA TCATGTGGGAGCA-3' and FLC3UTR 5'-CTCACACGAATAAGGTACAAA GTTC-3'. Labelling was performed with FLCexon7 primer in a primer extension reaction.

FLC antisense quantification was undertaken using qPCR with primers F 5'-TTTTTTTTTTTTTTTTTCTGCTTCCA-3' and R 5'-CACACCACCAAATA ACAACCA-3' known as set 1 and normalized against the Arabidopsis UBC gene²¹. The first strand cDNA was synthesized with an oligo(dT) primer.

FLC sense spliced (spliced exon 4/5, 5'-AGCCAAGAAGACCGAACTCA-3'; exon 7, 5'-TTTGTCCAGCAGGTGACATC-3') and FLC unspliced (unspliced exon 1, 5'-TTCTCCTCCGGCGATAAGTA-3'; unspliced intron1, 5'-TCACT CAACAACATCGAGCAC-3'; unspliced intron2, 5'-CGCAATTTTCATA GCCCTTG-3'; unspliced intron3, 5'-CTTTGTAATCAAAGGTGGAGAGC-3') quantification was achieved using a mixture of gene-specific-primer-primed cDNA (exon 7, unspliced intron1, unspliced intron3, UBC_R).

GFP mRNA quantification was achieved using the primers GFP_qPCR_LP, 5'-TGCAGTGCTTCTCCCGTTAC-3'; GFP_qPCR_RP, 5'-GGTCCTCTCCTG CACGTATC-3'. The GFP antisense transcript was detected by RT–PCR with primers GFPas2_F, 5'-TCAAGGACGACGGGAACTAC-3' (or GFPas1_F, 5'-GTCCACACAATCTGCCCTTT-3') and R, 5'-TTGACAGAAGTGAAGAACA CATACA-3' on a cDNA synthesized with GFPas2_F or GFPas1_F using Phusion (Finnzymes).

VIN3 levels were quantified using the primers VIN3_F, 5'-GTATGG GATTGGGAGTGATGAT-3' and VIN3_R, 5'-CAAAACAACCTGAAACCTG TGA-3'.

The *FLC* antisense RNA shown in Supplementary Fig. 2 was quantified by qPCR using primers F (5'-ACCTTATTCGTGTGAGAATTGC-3') and R (5'-TTGACAGAAGTGAAGAACACATACA-3'), (known as set 6) on cDNA synthesized using oligo(dT) and set6_F and normalized against *UBC*.

cDNA was synthesized using SuperScript III (Invitrogen) and analysed by quantitative PCR on an OPTICON2 instrument and using SYBR Green Jump-Start Taq ReadyMix from Sigma.

Constructs. pMDC43²² and the Gateway system (Invitrogen) was used to clone an *FLC* antisense promoter region (5'-CACCCTTCCGGTGACTCTCCCACTA-3' and 5'-TTTAACAAACCTTTCACTTT-3') or the equivalent region from the *Arabidopsis rbcS3A* gene (5'-CACCAACACCCGTCAAGTCCAA-3' and 5'-GGTGGGATCCAAAAATCAAG-3'). The resulting plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 and introduced into Col-0 FRI(Sf2) plants²³ by *Agrobacterium*-mediated floral dip transformation²⁴.

The COOLAIR promoter fusion construct was constructed by MultiSite Gateway (Invitrogen)-mediated recombination into a pRK290-based *Agrobacterium* binary vector. The fragments were amplified using the primers COOLAIRprom_F, 5'-GGGGACTGCTTTTTTGTACAAACTTGTCAATCTTC CGGTGACTCTCC-3' and COOLAIRprom_R, 5'-GGGGACAACTTTGTATA GAAAAGTTGTTTTTAACAAACCTTTCACTTT-3' on genomic DNA from Col-0 plants; BOX2_F, 5'-GGGGACAAGTTTGACAAAAAAGCAGGCTA CGACGACTCCTCCTGTA-3' and BOX2_R, 5'-GGGGACCACTTTGT ACAAGAAAAGCTGGGTTTGATCAATTCCACAGTTTTC-3' on pCambia13812²⁵; LUC_F, 5'-GGGGACAGCTTTCTTGTACAAAGTGGTGTAGATAGATAGCGATC CGGT-3' and LUC_R, 5'-GGGGACAACTTTGTATAAAAAGTTGGACCCGATC TAGTAACATAGA-3' on yy376 (ref. 26). The *FLC*-LUC translational fusion was described previously²⁷. The control plasmid was created by fusing a 1-kilobase sequence upstream from the ATG codon of At3g63530, a gene which expression is not affected by cold, into a pBar vector²⁸.

RACE. FLC antisense transcript mapping was performed as described in ref. 29 and the 5'RACE was performed using an Invitrogen RACE kit.

Chromatin immunoprecipitation. Chromatin immunoprecipitation and primers previously described in ref. 30 were used in combination with an RNA polII antibody 8WG16 from Abcam (ab817). Actin was used as an internal control using primers described in ref. 30. The average precipitation for actin of the

non-vernalized samples was 2.69% of input, compared to 2.33% of input in vernalized samples.

Microarray. RNA was extracted using the method described in ref. 31. rRNA was removed with a RiboMinus Plant Kit (Invitrogen). The resulting RNA was labelled using the Affymetrix GeneChip Whole Transcript Sense Target Labelling Assay. Probes were labelled and hybridized by Cogenics on custom design Affymetrix/Nimble Express GeneChips. The arrays contained 25-nucleotide oligonucleotides tiled with single nucleotide resolution across both strands of the 50 kb region of chromosome 5 including FLC. We also printed probes corresponding to genes used for expression normalization corresponding to oligonucleotides used on the Affymetrix ATH1 array and shown by ref. 21 to be optimal as reference genes for transcript normalization. In addition, we included genes we had used previously to normalize expression in northern blots and qPCR experiments (250317at, 244918at, 259361at, 253287at, 259407at, 262909at, 253355at, 265256at). The microarrays were analysed using the statistical analysis language R32; using the LIMMA33 and CNA34 libraries under the BioConductor frame work³⁵. We used an algorithm that computes a global rankinvariant set of transcripts; which was then in turn employed to normalize all microarrays across experimental conditions³⁵. Following normalization, linear models were fitted to the data according to ref. 33; and t-scores for all contrasts of interest were obtained. The t-scores were used as more reliable estimates of differential expression than averaged expression ratios as we let variation affect the quality of the scores. Segments of scores were obtained using the method of ref. 34; where P-values were obtained though permutations of t-scores contained in segments.

- 21. Czechowski, T. et al. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol*. **139**, 5–17 (2005).
- 22. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469 (2003).
- 23. Lee, I., Michaels, S. D., Masshardt, A. S. & Amasino, R. M. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of *Arabidopsis*. *Plant J.* **6**, 903–909 (1994).
- Bechtold, N., Ellis, J. & Pelletier, G. In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. C.R. Acad. Sci. III 316, 1194–1199 (1993)
- Hajdukiewicz, P., Svab, Z. & Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* 25, 989–994 (1994).
- Yamamoto, Y. Y. et al. Gene trapping of the Arabidopsis genome with a firefly luciferase reporter. Plant J. 35, 273–283 (2003).
- Mylne, J. S. et al. LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. Proc. Natl Acad. Sci. USA 103, 5012–5017 (2006).
- Völker, A., Stierhof, Y. D. & Jürgens, G. Cell cycle-independent expression of the *Arabidopsis* cytokinesis-specific syntaxin KNOLLE results in mistargeting to the plasma membrane and is not sufficient for cytokinesis. J. Cell Sci. 114, 3001–3012 (2001)
- Liu, F. et al. The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. Mol. Cell 28, 398–407 (2007).
- De Lucia, F. et al. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc. Natl Acad. Sci. USA 105, 16831–16836 (2008).
- Quesada, V., Macknight, R., Dean, C. & Simpson, G. G. Autoregulation of FCA premRNA processing controls Arabidopsis flowering time. EMBO J. 22, 3142–3152 (2003).
- 32. R. Development Core Team. R: a language and environment for statistical computing. (http://www.R-project.org) (R Foundation for Statistical Computing, 2006)
- Smyth, G. K. in Bioinformatics and Computational Biology Solutions using R and Bioconductor (eds Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer, 2005).
- 34. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557–572 (2004).
- 35. Pelz, C. R., Kulesz-Martin, M., Bagby, G. & Sears, R. C. Global rank-invariant set normalization (GRSN) to reduce systematic distortions in microarray data. *BMC Bioinformatics* **9**, 520 (2008).

LETTERS

Long-range oncogenic activation of *Igh-c-myc* translocations by the *Igh* 3' regulatory region

Monica Gostissa^{1,2,3,4}, Catherine T. Yan^{1,2,3,4}, Julia M. Bianco^{1,2,3,4}, Michel Cogné⁵, Eric Pinaud⁵ & Frederick W. Alt^{1,2,3,4}

B-cell malignancies, such as human Burkitt's lymphoma, often contain translocations that link c-myc or other proto-oncogenes to the immunoglobulin heavy chain locus (IgH, encoded by Igh)1. The nature of elements that activate oncogenes within such translocations has been a long-standing question. Translocations within Igh involve DNA double-strand breaks initiated either by the RAG1/2 endonuclease during variable, diversity and joining gene segment (V(D)J) recombination, or by activation-induced cytidine deaminase (AID, also known as AICDA) during class switch recombination (CSR)²⁻⁴. V(D)J recombination in progenitor B (pro-B) cells assembles Igh variable region exons upstream of μ constant region (Cµ) exons, which are the first of several sets of C_H exons ('C_H genes') within a C_H locus that span several hundred kilobases (kb)^{5,6}. In mature B cells, CSR deletes Cμ and replaces it with a downstream C_H gene⁶. An intronic enhancer ($iE\mu$) between the variable region exons and Cµ promotes V(D)J recombination in developing B cells⁷. Furthermore, the Igh 3' regulatory region (Igh3' RR) lies downstream of the C_H locus and modulates CSR by long-range transcriptional enhancement of C_H genes⁸⁻¹⁰. Transgenic mice bearing iEu or Igh3'RR sequences fused to c-myc are predisposed to B lymphomas, demonstrating that such elements can confer oncogenic c-myc expression¹¹⁻¹⁶. However, in many B-cell lymphomas, Igh-c-myc translocations delete iEu and place c-myc up to 200 kb upstream of the Igh3' RR1. Here we address the oncogenic role of the Igh3' RR by inactivating it in two distinct mouse models for B-cell lymphoma with Igh-c-myc translocations. We show that the Igh3' RR is dispensable for pro-B-cell lymphomas with V(D)J recombination-initiated translocations, but is required for peripheral B-cell lymphomas with CSR-associated translocations. As the Igh3'RR is not required for CSR-associated Igh breaks or Igh-c-myc translocations in peripheral B-cell lymphoma progenitors, we conclude that this regulatory region confers oncogenic activity by long-range and developmental stage-specific activation of translocated c-myc genes.

Individual C_H genes are organized into germline transcription units which consist, from 5' to 3', of a non-coding 'I' exon, a switch (S) region and the C_H exons⁶. CSR to a particular C_H gene requires the introduction of AID-initiated DNA double-strand breaks (DSBs) into the donor S region upstream of C_H (S $_H$) and into a downstream acceptor S region⁶. The I exon is preceded by a germline promoter that is upregulated by particular activation treatments, with transcription targeting AID to specific S regions¹⁷. The Igh3'RR contains multiple enhancer elements^{18–20} and controls germline transcription of certain C_H promoters over distances of 100 kb or more⁸. However, the Igh3'RR is not required for V(D)J recombination, expression of rearranged Igh genes, or transcription through S $_H$ or S $_H$ 1 (ref. 10). To test the potential roles of the Igh3'RR in B-cell lymphomagenesis, we

bred an *Igh3' RR* inactivating mutation, which deletes the key hs3b and hs4 enhancers¹⁰ (Supplementary Fig. 1), into non-homologous endjoining (NHEJ) and p53 tumour suppressor deficient backgrounds that predispose to either pro-B or peripheral B-cell lymphoma.

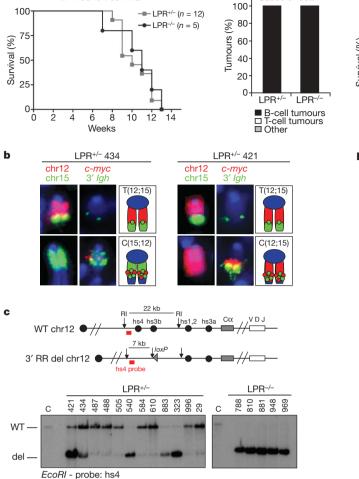
XRCC4 and DNA ligase 4 (Lig4) form a NHEJ ligation complex required for V(D)J recombination²¹. Mice with germline inactivation of either Lig4 or Xrcc4 and the p53 tumour suppressor develop pro-Bcell lymphomas with complex translocations ('complicons') involving Igh on chromosome 12 (chr12) and a region downstream of *c-myc* on chromosome 15 $(chr15)^{22}$. These translocations arise from joining RAG1/2-induced DSBs in the Igh JH region to DSBs downstream of c-myc, leading to dicentric chr12;15 translocations and c-myc amplification by breakage-fusion-bridge cycles^{22,23}. Analysis of Lig4 and p53 double-deficient mice ('LP' mice) that contained Igh3'RR inactivating mutations on either one or both Igh alleles (referred to as LPR^{+/-} or LPR^{-/-}, respectively) showed that both genotypes succumbed to pro-B-cell lymphomas with kinetics similar to those of LP mice²² (Fig. 1a and Supplementary Table 1). Likewise, all analysed LPR^{+/-} and LPR^{-/-} tumours had characteristic chr12 to chr15 translocations (T(12;15)) and 15;12 complicons (C(15;12)) containing amplified c-myc (Fig. 1b, Supplementary Fig. 2 and Supplementary Table 1). Furthermore, Southern blotting with a probe that distinguishes Igh3' RR-deleted from wild-type Igh alleles (Fig. 1c, top) showed that Igh-c-myc translocations/amplifications involved the wild-type allele in some LPR^{+/-} tumours and the *Igh3' RR*-deleted allele in others (Fig. 1c, bottom). Thus, the Igh3' RR is dispensable for LP pro-B-cell lymphomas with *Igh–c-myc* complicons.

Specific inactivation of *Xrcc4* by a *loxP*/Cre approach in peripheral B cells of p53-deficient mice (referred to as 'CXP' mice) leads to surface immunoglobulin-negative peripheral B-cell lymphomas^{24,25}. CXP B-cell lymphomas arise from progenitors that delete or aberrantly rearrange their Igk and Igl light-chain loci and routinely contain a T(12;15) that fuses Igh S regions to sequences just upstream of c-myc, leading to high level c-myc expression from a translocated single copy c-myc gene^{25,26}. Such T(12;15)s lack $iE\mu$, as they occur downstream of this element²⁴. To test potential roles of the *Igh3' RR* in CXP tumorigenesis, we followed tumour development in cohorts of CXPR^{+/-} and CXPR^{-/-} mice. CXPR^{+/-} mice succumbed to the same tumour spectrum as CXP mice²⁵, with 40% developing surface immunoglobulin-negative B-cell lymphomas that appeared identical to CXP B-cell lymphomas (Fig. 2a and Supplementary Table 2). The remaining CXPR+/- mice succumbed to thymic lymphomas or other tumours associated with germline p53 deficiency. In contrast, none of 17 analysed CXPR^{-/-} mice developed B-cell lymphoma; instead, most died from thymic lymphoma (Fig. 2a and Supplementary Table 2). Thus, homozygous Igh3'RR inactivation abrogates CXP B-cell lymphomas.

¹Howard Hughes Medical Institute, ²The Children's Hospital, ³Immune Disease Institute, ⁴Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵Centre National de la Recherche Scientifique; Université de Limoges, UMR 6101, Faculté de Médecine, Limoges France.

LPR cohort survival

a



Cause of death

Figure 1 | Deletion of the Igh3'RR does not affect development of pro-B-cell **lymphomas.** a, Left, Kaplan–Meier curve of the LPR^{+/-} (n = 12) and LPR $^{-/-}$ (n = 5) cohorts. Curves represent total survival. Right, the percentage of mice in the LPR^{+/-} and LPR^{-/-} cohorts succumbing to B-cell lymphomas, thymic lymphomas or other causes of death. $\bf b$, Examples of cytological aberrations in representative LPR $^{+/-}$ tumours with translocations to the wild-type (mouse number 434) or the 3'RR-deleted (mouse number 421) Igh alleles. In each set of panels: left, paints specific for chr12 (red) and chr15 (green); middle, FISH analysis on separated metaphases with 3' Igh (green) and c-myc (red) probes; right, graphic representation. Only chromosomes involved in translocations are shown. Whole metaphases are presented in Supplementary Fig. 2. c, Southern blot analysis of LPR^{+/-} (left) and LPR^{-/-} (right) tumour DNA with a probe downstream of hs4, which distinguishes the wild-type (WT) and 3'RRdeleted (del) Igh alleles. A schematic of the wild-type and del Igh locus, with the position of the probe, is on the top. Numbers refer to individual mice in the cohort (see Supplementary Table 1). C, control, total spleen DNA from wild-type mouse; RI, EcoRI.

All analysed CXPR^{+/-} B-cell tumours had a clonal T(12;15), on the basis of spectral karyotyping (Fig. 3c and Supplementary Fig. 3). As in CXP B-cell lymphomas²⁵, most T(12;15) from CXPR^{+/-} B-cell lymphomas split *c-myc*, as evidenced by Southern blotting with 5' and 3' *c-myc* locus probes (Fig. 2b, top). However, two tumours (393 and 959) had *c-myc* amplification without detectable *c-myc* rearrangements (Fig. 2b), and tumour 796, while harbouring a clonal T(12;15) that split *c-myc*, also contained metaphases with a translocation that fused this T(12;15) to chromosome 16, resulting in low level *Igh/c-myc* amplification (Supplementary Fig. 3 and data not shown). Northern blotting showed increased *c-myc* expression in all CXPR^{+/-} B-cell tumours, with tumours 393, 796 and 959 showing the highest levels (Supplementary Fig. 4), suggesting that amplification may sometimes be selected secondary to ectopic activation to achieve maximal

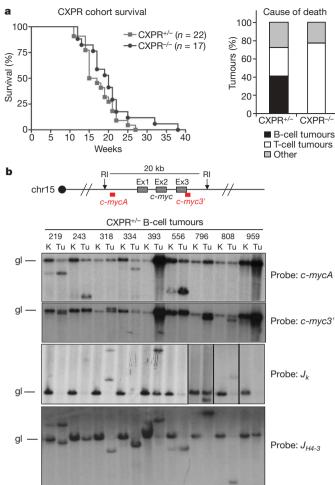


Figure 2 | Deletion of the *Igh3'RR* abrogates development of peripheral B-cell lymphomas. a, Left, Kaplan–Meier curve of the CXPR $^{+/-}$ (n=22) and CXPR $^{-/-}$ (n=17) cohorts. Curves represent total survival. Right, the percentage of mice in the CXPR $^{+/-}$ and CXPR $^{-/-}$ cohorts succumbing to B-cell lymphomas, thymic lymphomas or other cause of death. b, Southern blot analysis of CXPR $^{+/-}$ tumour DNA with probes indicated. Numbers refer to individual mice in the cohort (see Supplementary Table 2). gl, germ line; K, kidney, used as control; Tu, tumour. A schematic of the *c-myc* locus, indicating the 5' and 3' probes used to detect *c-myc* rearrangements is on the top.

expression, which may be relevant to certain human lymphomas that acquire *c-myc* amplification during tumour progression¹².

To determine which *Igh* allele was involved in CXPR^{+/-} lymphoma T(12;15)s, we analysed tumour metaphases by fluorescence *in situ* hybridization (FISH) with a probe specific for the deleted portion of the *Igh3'RR* (3'RR wild-type probe, green) and a chr15 paint (red). In this assay, the translocated portion of chr15 (red paint) co-localizes with a green signal if the wild-type *Igh* allele is translocated, but not if the *Igh3'RR*-deleted allele is translocated (Fig. 3a, left). Sequential re-probing of these metaphases with a green chr12 paint (2nd FISH, green) plus a 3' *Igh* BAC probe (2nd FISH, red) (Fig. 3a, right) showed that all CXPR^{+/-} tumour *Igh-c-myc* translocations involved the wild-type *Igh* allele (Fig. 3b, c and Supplementary Fig. 5). These results, coupled with the absence of B-cell tumours in CXPR^{-/-} mice, demonstrate that the *Igh3'RR* is required for peripheral CXP B-cell lymphomas by a role in oncogenic *Igh-c-myc* translocations.

The *Igh3' RR* might influence the appearance of oncogenic *Igh–c-myc* translocations by mechanistically promoting them through induction of CSR at certain S regions and/or by long-range activation of translocated *c-myc* expression. To distinguish between these

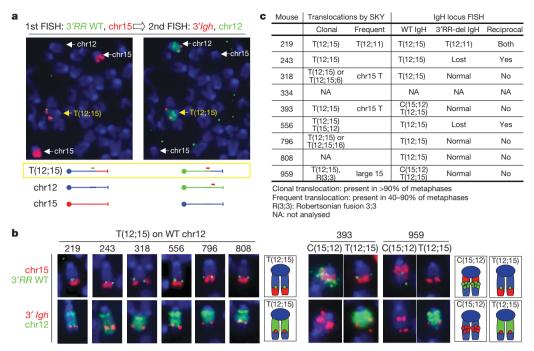


Figure 3 | T(12;15) in CXPR^{+/-} B-cell tumours always involves the wildtype *lgh* allele. a, Example of FISH and chromosome paints on CXPR^{+/-} tumour metaphases. The same metaphase was sequentially analysed with the different set of probes indicated at the top of each panel. A schematic of the different chromosomal species detected is shown at the bottom. b, Summary of FISH and chromosome paint analyses on all analysed CXPR^{+/-} tumours.

Numbers refer to individual mice in the cohort. Sequential hybridization with the set of probes indicated on the left was performed. Only chromosomes involved in translocations are shown, along with a graphic representation. The whole metaphases are presented in Supplementary Fig. 5. **c**, Table summarizing spectral karyotyping (SKY) and FISH data for CXPR^{+/-} B-cell tumours.

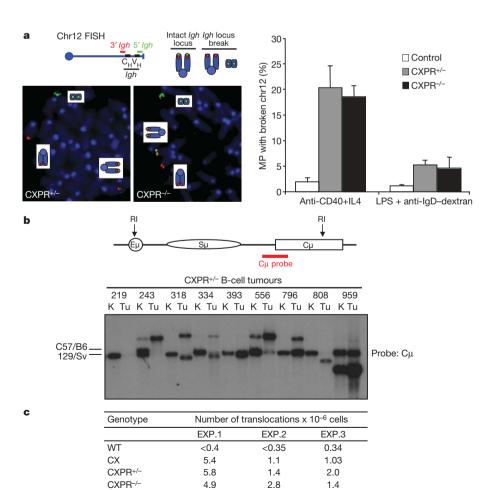


Figure 4 | Deletion of the Iah3'RR does not affect Igh locus breaks and Igh-c-myc translocations. a, Left, top: diagram showing location of 3' Igh (red) and 5' Igh (green) BAC probes used for FISH. An intact Igh locus on chr12 shows colocalized red and green signals, a broken locus shows split red and green signals. Left, bottom: examples of metaphases from CXPR^{+/-} and CXPR^{-/-} splenic B cells stimulated with anti-CD40 and IL4, showing Igh breaks. Right: quantification of *Igh* locus breaks in B cells activated for 4 days with anti-CD40 and IL4 or for 5 days with LPS and anti-IgD-dextran, from control, CXPR^{+/-} and CXPR⁻ mice. At least three mice per each genotype and at least 60 metaphases per mouse were analysed; data are presented as mean and s.d. b, Southern blot analysis of CXPR^{+/-} tumour DNA with a Cμ probe that detects Sµ rearrangements. A schematic of the *Igh* locus, indicating the probe position, is on the top. The position of germline bands in C57/B6 and 129/Sv backgrounds is indicated on the left; the 3'RR-deleted allele is from 129/Sv background. Numbers refer to individual mice in the cohort. Note that in some cases kidneys contained infiltration of tumour cells, as judged by tumour-specific rearranged J_H and c-myc bands (Fig. 2b). c, Frequency of Igh-cmyc translocations was measured by PCR assays using Sµ and *c-myc* primers. DNA samples were isolated from day 4 splenic B cells from wild-type, CX, CXPR^{+/-} and CXPR^{-/-} mice, activated with anti-CD40 and IL4.

potential mechanisms, we asked whether the Igh3' RR-deleted allele is a substrate for AID-induced DSBs. As one measure, we activated and CXPR^{-/-} B cells for 4 days with anti-CD40 and IL4 and found that both genotypes switched to IgG1, as expected by the fact that the 3' IghRR is not required for CSR to this IgH isotype^{8,10} (Supplementary Fig. 6a). We also analysed metaphases from anti-CD40- and IL4-stimulated CXPR^{+/-} and CXPR^{-/-} B cells using two-colour FISH²⁷ and found that the increased level of *Igh* breaks in the absence of XRCC4 (ref. 24) was not markedly affected by the Igh3' RR mutation (Fig. 4a). CSR to IgG3 is severely impaired in mice homozygous for Igh3' RR inactivating mutations due to inhibition of Iγ3 transcription^{8,10}. However, lipopolysaccharide (LPS) and anti-IgD-dextran stimulation to induce IgG3 CSR led to similarly increased levels of Igh breaks in CXPR+/- and CXPR-/- B cells (Supplementary Fig. 6b and Fig. 4a), probably reflecting unimpeded AID activity on Su, which is transcribed independently of the Igh3' RR¹⁰. Southern blotting further showed that most CXPR^{+/-} B-cell tumours had Su rearrangements or deletions on both alleles, again indicative of substantial AID activity on Igh3' RR-deleted alleles (Fig. 4b). We conclude that the introduction of AID-induced Igh lesions is not markedly impaired by the *Igh3' RR* deletion.

Although AID-induced *Igh* breaks occur at high frequency on *Igh3' RR*-deleted alleles, *Igh-c-myc* translocations still might be inhibited, for example by a different distribution of *Igh* DSBs or by effects on proximity of the two loci²⁶. Therefore, we used a PCR assay²⁸ to directly evaluate the potential effects of the *Igh3' RR* deletion on *Igh-c-myc* translocation frequency, and found that *Igh-c-myc* translocations, indeed, occurred at similar frequencies in CXPR^{+/-} and CXPR^{-/-} B cells stimulated with anti-CD40 and IL4 (Fig. 4c). Moreover, one CXPR^{+/-} B-cell tumour had a T(12;11) involving the *Igh3' RR*-deleted allele in addition to its T(12;15), again indicating the *Igh3' RR*-deleted allele is a translocation target (Fig. 3c and Supplementary Fig. 3). We conclude that the *Igh3' RR* is dispensable for generation of *Igh-c-myc* translocations in XRCC4-deficient B cells.

The *Igh* locus has long been speculated to have *cis*-acting elements that activate c-myc or other oncogenes in the context of translocations. We now demonstrate that the Igh3' RR is required for oncogenicity of *Igh–c-myc* translocations that ectopically activate *c-myc* in mouse CXP B-cell lymphomas. As a substantial proportion of CXPR^{+/-} (Supplementary Fig. 7) and CXP lymphomas²⁵ have translocations that fuse *c-myc* to Sμ, oncogenic *Igh3' RR* activity extends at least 200 kb upstream. Thus, our findings define a major oncogenic role for the Igh3' RR in activating c-myc subsequent to Igh-c-myc translocations; although we do not rule out an additional role in promoting translocations by enhancing AID-mediated lesions in certain S regions regulated by this element. Although high-copy $iE\mu$ transgenes predispose to pro-B-cell lymphoma in mice¹², the role of $iE\mu$ in Igh–oncogene translocations remains to be determined. In this context, knock-in of *c-myc* into the *Igh* J_H region led to peripheral B-cell lymphomas, as opposed to pro-B-cell lymphomas¹², suggesting that $iE\mu$ alone may not always be sufficient to activate *c-myc* in the endogenous setting. In this regard, our finding that the Igh3' RR is dispensable for LP pro-B-cell lymphomas may explain why *c-myc* is amplified in these tumours and ectopically activated in peripheral CXP B-cell tumours. Specifically, the Igh3' RR is not active in pro-B cells9 and would not activate Igh translocations upstream of a singlecopy c-myc, favouring selection for translocations downstream of c-myc that promote gene amplification. Given the similar organization of mouse and human Igh9, our findings suggest that the Igh3' RR also supports activated oncogene expression in human B-cell tumours with Igh S region translocations that eliminate $iE\mu$ (for example, sporadic Burkitt's lymphoma²⁹) and, potentially, even in some with Igh J_H region translocations that leave $iE\mu$ intact (for example, endemic Burkitt's lymphoma²⁹). In this regard, targeted inhibition of the B-cell-specific Igh3' RR could theoretically provide a therapeutic strategy for such human B lymphomas.

METHODS SUMMARY

Mouse strains. *Igh3'RR*-deleted mice were generated previously ¹⁰ and crossed into $Lig4^{+/-}p53^{+/-}$ (ref. 30) or *CD21-Cre, Xrcc4^{c/c}p53^{+/-}* (ref. 25) mice to obtain triple or quadruple heterozygous animals, which were appropriately crossed to obtain the experimental cohorts. Mice were analysed as outlined in the text at 8–30 weeks of age. The Institutional Animal Care and Use Committee of Children's Hospital (Boston, Massachusetts) approved all animal work.

Splenic B-cell purification, activation in culture and CSR assays. $CD43^-$ B cells were isolated and cultured as previously described²⁷. Cells were processed at day 4 of stimulation with anti-CD40 and IL4, or at day 5 of stimulation with LPS and anti-IgD–dextran for DNA isolation, metaphase preparation and flow cytometry analysis (see Methods).

Two-colour FISH. Metaphase spreads were prepared and FISH experiments performed according to standard protocols²⁷. FISH probes are detailed in Methods. Whole chromosome paints specific for mouse chromosome 12 and 15 were used according to the manufacturer's instructions (Applied Spectral Imaging).

PCR assay to detect *Igh–c-myc* translocations. *Igh–c-myc* translocation junctions were amplified by PCR from genomic DNA prepared from splenic B cells activated for 4 days with anti-CD40 and IL4, using primers and conditions previously described²⁸ (see Methods). DNA corresponding to 50,000 or 100,000 cells was analysed in separate reactions. PCR products were run on agarose gel, blotted and hybridized with an internal oligonucleotide probe in the *c-myc* locus.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 September; accepted 30 October 2009.

- Küppers, R. & Dalla-Favera, R. Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene 20, 5580–5594 (2001).
- Lieber, M. R., Yu, K. & Raghavan, S. C. Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations. *DNA Repair (Amst.)* 5, 1234–1245 (2006).
- Ramiro, A. et al. The role of activation-induced deaminase in antibody diversification and chromosome translocations. Adv. Immunol. 94, 75–107 (2007).
- Robbiani, D. F. et al. AID is required for the chromosomal breaks in c-myc that lead to c-myc/lgH translocations. Cell 135, 1028–1038 (2008).
- Bassing, C. H., Swat, W. & Alt, F. W. The mechanism and regulation of chromosomal V(D)J recombination. Cell 109 (suppl.), S45–S55 (2002).
- Chaudhuri, J. et al. Evolution of the immunoglobulin heavy chain class switch recombination mechanism. Adv. Immunol. 94, 157–214 (2007).
- Henderson, A. & Calame, K. Transcriptional regulation during B cell development. Annu. Rev. Immunol. 16, 163–200 (1998).
- 8. Cogné, M. et al. A class switch control region at the 3' end of the immunoglobulin heavy chain locus. *Cell* 77, 737–747 (1994).
- Khamlichi, A. A., Pinaud, E., Decourt, C., Chauveau, C. & Cogne, M. The 3' IgH regulatory region: a complex structure in a search for a function. Adv. Immunol. 75, 317–345 (2000).
- 10. Pinaud, E. *et al.* Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* 15, 187–199 (2001).
- Adams, J. M. et al. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. Nature 318, 533–538 (1985).
- 12. Janz, S. Myc translocations in B cell and plasma cell neoplasms. DNA Repair (Amst.) 5, 1213–1224 (2006).
- Schmidt, E. V., Pattengale, P. K., Weir, L. & Leder, P. Transgenic mice bearing the human c-myc gene activated by an immunoglobulin enhancer: a pre-B-cell lymphoma model. *Proc. Natl Acad. Sci. USA* 85, 6047–6051 (1988).
- 14. Truffinet, V. et al. The 3' IgH locus control region is sufficient to deregulate a c-myc transgene and promote mature B cell malignancies with a predominant Burkitt-like phenotype. J. Immunol. 179, 6033–6042 (2007).
- Wang, J. & Boxer, L. M. Regulatory elements in the immunoglobulin heavy chain gene 3'-enhancers induce c-myc deregulation and lymphomagenesis in murine B cells. J. Biol. Chem. 280, 12766–12773 (2005).
- Yan, Y., Park, S. S., Janz, S. & Eckhardt, L. A. In a model of immunoglobulin heavychain (*IGH*)/MYC translocation, the *Igh* 3' regulatory region induces MYC expression at the immature stage of B cell development. *Genes Chromosom.* Cancer 46, 950–959 (2007).
- Manis, J. P., Tian, M. & Alt, F. W. Mechanism and control of class-switch recombination. *Trends Immunol.* 23, 31–39 (2002).
- Madisen, L. & Groudine, M. Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates c-myc expression in plasmacytoma and Burkitt's lymphoma cells. Genes Dev. 8, 2212–2226 (1994).
- Dariavach, P., Williams, G. T., Campbell, K., Pettersson, S. & Neuberger, M. S. The mouse IgH 3'-enhancer. Eur. J. Immunol. 21, 1499–1504 (1991).
- Lieberson, R., Giannini, S. L., Birshtein, B. K. & Eckhardt, L. A. An enhancer at the 3' end of the mouse immunoglobulin heavy chain locus. *Nucleic Acids Res.* 19, 933–937 (1991).

- Lieber, M. R. The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. 283, 1–5 (2008).
- Zhu, C. et al. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. Cell 109, 811-821 (2002).
- Difilippantonio, M. J. et al. Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. J. Exp. Med. 196, 469–480 (2002).
- 24. Yan, C. T. et al. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
- Wang, J. H. et al. Oncogenic transformation in the absence of Xrcc4 targets peripheral B cells that have undergone editing and switching. J. Exp. Med. 205, 3079–3090 (2008).
- Wang, J. H. et al. Mechanisms promoting translocations in editing and switching peripheral B cells. Nature 460, 231–236 (2009).
- 27. Franco, S. et al. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. Mol. Cell 21, 201–214 (2006).
- Ramiro, A. R. et al. Role of genomic instability and p53 in AID-induced c-myc-lgh translocations. Nature 440, 105–109 (2006).
- Neri, A., Barriga, F., Knowles, D. M., Magrath, I. T. & Dalla-Favera, R. Different regions of the immunoglobulin heavy-chain locus are involved in chromosomal translocations in distinct pathogenetic forms of Burkitt lymphoma. *Proc. Natl Acad. Sci. USA* 85, 2748–2752 (1988).

30. Frank, K. M. et al. DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. Mol. Cell 5, 993–1002 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. G. Jiang for help with cloning translocation breakpoints. This work was supported by National Institutes of Health (NIH) grant CA92625 and the Leukemia and Lymphoma Society of America (LLS) SCORE grant to F.W.A. M.G. was a Special Fellow of the LLS and C.T.Y. was supported by an NCI training grant. M.G. and C.T.Y. were also supported by the de Villiers International Achievement Award of the LLS to F.W.A. E.P. was a Fellow of the Fondation pour la Recherche Médicale. F.W.A. is an Investigator of the Howard Hughes Medical Institute

Author Contributions F.W.A. and M.G. planned studies and interpreted data. M.G. performed experiments, with technical help from J.M.B. C.T.Y generated CXP mice. E.P. and M.C. generated *lgh3'RR*-deleted mice and helped interpret data. F.W.A. and M.G. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu).

doi:10.1038/nature08633 nature

METHODS

Southern blotting. Genomic DNA was separated on 0.8% agarose gel and transferred to Zeta-Probe GT (Biorad) nylon membrane. Hybridization was performed in 50% formamide/SScPE at 42 °C. The J_{H4-3} probe was a 1.6-kb HindIII/EcoRI fragment downstream of J_{H4} ; the J_K probe was a 1-kb BgIII–HindIII fragment downstream of J_{K5} ; c-mycA probe was a 1.7-kb XbaI–XbaI, genomic fragment upstream of c-myc exon 1; c-myc3′ probe was a 1.4-kb XhoI–KpnI genomic fragment that contains part of exon 3 and the 3′ untranslated of c-myc.

Splenic B-cell purification, activation in culture and CSR assays. B cells were isolated by negative selection using CD43 MACS MicroBeads and LS columns (Miltenyi). Cells were put in culture at the density of 5×10^5 per ml and cultured for 4 days in the presence of anti-CD40 $(1\,\mu\mathrm{g\,ml}^{-1},\ eBioscience)$ and IL4 $(20\,\mathrm{ng\,ml}^{-1},\ R\&D\ Systems)$ or for 5 days in the presence of LPS (25 ng ml $^{-1},\ Sigma)$ and anti-IgD–dextran $(3\,\mathrm{ng\,ml}^{-1},\ Fina\ BioSolutions\ LLC)$. CSR was assessed by staining the cells with CyChrome (CyC)-labelled anti-mouse B220 (eBiosciences) and phycoerythrin (PE)-labelled anti-mouse IgG1 (BD Biosciences) or FITC-labelled anti-mouse IgG3 (BD Biosciences).

Two-colour FISH and SKY. For metaphase preparation, cells were incubated for 3 h with colcemid (KaryoMAX, GIBCO), swollen in 75 mM potassium chloride for 25 min at 37 °C, fixed in methanol/acetic acid (3/1), and air dried on slides overnight. The following BACs were used as probes for FISH: BAC199, covering the 3′ region of the *Igh* locus encompassing the *Igh3′RR* and 100 kb downstream (3′ *Igh* BAC); BAC207, upstream of the *Igh* V_H region (5′ *Igh* BAC), and BAC RP23-307D14, containing the *c-myc* locus (*c-myc* BAC). The 3′RR wild-type probe was prepared by PCR-amplifying a DNA fragment containing the hs3b and hs4 region using the following primers: 5′-TGGGCTTGCTTTGTTCTC CATTCT-3′ and 5′-AGGTCTACACAGGGGCTCTG-3′. BAC DNA was labelled with either DIG-11-dUTP or biotin-16-dUTP (Roche) according to standard protocols. Probes were visualized with avidin-Cy3 or antidigoxigenin-FITC (Roche, 1:400 dilution). Whole chromosome paints specific for mouse chromosome 12 and 15 were used according to the manufacturer's instructions (Applied Spectral Imaging). Images were captured with a ×60 objective using

the BX61 microscope (Olympus) equipped with an automatic stage, a cooled-CCD camera and an interferometer (Applied Spectral Imaging). Analysis was performed with the ScanView software (Applied Spectral Imaging). For sequential hybridizations, slides were first incubated with the 3'RR wild-type probe (visualized in green) plus chr15 paint (visualized in red); at least 15 metaphases per tumour samples were acquired. Slides were then washed in 2× SSC, dehydrated through an ethanol series (70%-80%-100%) to strip the first set of probes, and reprobed with the 3' Igh probe (visualized in red) plus chr12 paint (visualized in green). The same metaphases analysed with the first set of probes were acquired again, using the automated stage with relocalization software. Stripping and reprobing was repeated once more, using the *c-myc* probe (visualized in red), plus the 5' Igh probe (visualized in green). Results from the second and third hybridizations were also confirmed on independent slides. Spectral karyotyping was performed with a mouse SKY paint kit (Applied Spectral Imaging) following manufacturer's indications. At least ten metaphases were analysed per each tumour sample. Analysis was performed with the HiSKY software (Applied Spectral Imaging).

PCR assay to detect *Igh–c-myc* translocations. For the first PCR rounds, the following primers were used: *c-myc* 3'a, 5'-GGGAAGGGGTGTCAAAT AATAAGA-3'; *Sm* 3'a, 5'-TGAGGACCAGAGAGGGATAAAAGAGA-3'. Conditions were as follows: 94 °C 3 min; 94 °C 15 s, 62 °C 15 s, 68 °C 7 min + 20 s per cycle, 25 cycles; 68 °C 5 min. For the nested PCR rounds, the following primers were used: *c-myc* 3'b, 5'-GACACCTCCCTTCTACACTCTAAACCG-3'; *Sm* 3'b, 5'-CACCCTGCTATTTCCTTGTTGCTAC-3'. Conditions were as follows: 94 °C 3 min; 94 °C 15 s, 62 °C 15 s, 68 °C 7 min, 25 cycles; 68 °C 5 min. The Expand Long Template PCR System (Roche) was used for both reactions.

Cloning of *Igh–c-myc* translocation breakpoints. Five-hundred nanograms of DNA from each tumour was used for amplification with the following primers: 5' *c-myc*R, 5'-AATCAAGGCGCTAGACGCGAGAAT-3'; $S\mu$ 3'a, 5'-TGAGG ACCAGAGAGGGATAAAAGAGAA-3'. Conditions were as follows: 94 °C 3 min; 94 °C 15 s, 58 °C 15 s, 68 °C 7 min, 35 cycles; 68 °C 5 min. The Expand Long Template PCR System (Roche) was used.

LETTERS

Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma

Min Guo¹, Yeeting E. Chong¹, Ryan Shapiro¹, Kirk Beebe¹†, Xiang-Lei Yang¹ & Paul Schimmel¹

Mistranslation arising from confusion of serine for alanine by alanyl-tRNA synthetases (AlaRSs) has profound functional consequences¹⁻³. Throughout evolution, two editing checkpoints prevent disease-causing mistranslation from confusing glycine or serine for alanine at the active site of AlaRS. In both bacteria and mice, Ser poses a bigger challenge than Gly^{1,2}. One checkpoint is the AlaRS editing centre, and the other is from widely distributed AlaXps—free-standing, genome-encoded editing proteins that clear Ser-tRNA^{Ala}. The paradox of misincorporating both a smaller (glycine) and a larger (serine) amino acid suggests a deep conflict for nature-designed AlaRS. Here we show the chemical basis for this conflict. Nine crystal structures, together with kinetic and mutational analysis, provided snapshots of adenylate formation for each amino acid. An inherent dilemma is posed by constraints of a structural design that pins down the α-amino group of the bound amino acid by using an acidic residue. This design, dating back more than 3 billion years, creates a serendipitous interaction with the serine OH that is difficult to avoid. Apparently because no better architecture for the recognition of alanine could be found, the serine misactivation problem was solved through free-standing AlaXps, which appeared contemporaneously with early AlaRSs. The results reveal unconventional problems and solutions arising from the historical design of the protein synthesis machinery.

The 20 aminoacyl-tRNA synthetases (AARSs; one for each amino acid) have a fundamental role in establishing faithful translation, through close control over the two-step aminoacylation reaction that establishes the genetic-code linkage between an amino acid and its cognate nucleotide triplet of the tRNA anticodon^{4,5}. Even a mild deviation from the rigorous accuracy of the relationship between nucleotide triplet and amino acid, caused by misaminoacylation, is toxic to bacteria and gives rise to serious pathology in the mouse^{1,2}. By sequentially working through two 'sieves'—first at the site for amino-acid activation and then at the centre for editing⁶⁻⁸—overall amino-acid specificity is the product of the discrimination achieved at each step9. For example, IleRS misactivates Val to produce ValtRNA^{Ile}. When the valyl moiety is translocated to the centre for editing, it fits snugly into a pocket that excludes the sterically larger Ile side chain¹⁰. Thus Val-tRNA^{IIe}, but not Ile-tRNA^{IIe}, is cleared¹¹. Conversely, ValRS cannot activate Ile but misactivates the isosteric Thr and uses a separate editing site to clear Thr-tRNA^{Val} (ref. 12). Paradoxically, this 'steric exclusion' mechanism does not operate for AlaRS, which activates the smaller (than Ala) glycine but also activates the sterically larger serine. This misactivation is seen in AlaRSs from bacteria to human^{2,13}.

Two mechanisms correct the misaminoacylation of Gly or Ser for Ala. One is associated with the editing centre of AlaRS¹. The second comes from separate, genome-encoded proteins (AlaXps) that are homologues of the editing domain of AlaRSs³,¹⁴. Thus, if misacylated

Ser-tRNA^{Ala} is produced, it is cleared by the editing centre of AlaRS or, failing that, it can be cleared by an AlaXp¹⁵—either before translocation to the ribosome or resampled from it¹⁶. AlaRS is the only tRNA synthetase that has a widely distributed (through all three kingdoms of life) homologue of its editing domain that is separately encoded by the genome^{14,17}. This observation suggests that overcoming the confusion of Gly or Ser for Ala was a significant challenge for nature. Toxicity from an editing defect in AlaRSs in both mouse and bacteria seems to be more problematic with Ser than with Gly^{1,2}. Consistent with these observations, whereas AlaRSs edit both mischarged Gly-tRNA^{Ala} and Ser-tRNA^{Ala} (ref. 1), all investigated AlaXps can hydrolyse Ser-tRNA^{Ala} but only some can hydrolyse Gly-tRNA^{Ala} (ref. 18) (Supplementary Fig. 2). It is plausible that the additional AlaXp-based editing mechanism developed to address the special challenge of serine for mistranslation.

Because serine is not simply bound but is activated by AlaRS to form seryl adenylate, we set out to obtain co-crystals of the three aminoacyl adenylates with the enzyme (see comment on previous work in Supplementary Information). The monomeric aminoterminal aminoacylation domain (residues 1-441) of Escherichia coli AlaRS is responsible for the synthesis of alanyl-tRNA^{Ala}. It is clearly separated from the carboxy-terminal half that is homologous to the freestanding editing protein AlaXp-II. Pursuant to obtaining an effective interface that could enhance the assembly of a stable crystal lattice, a surface leucine-zipper mutation—H104L/Q108L/E112L was introduced into AlaRS₄₄₁, the N-terminal catalytic fragment of E. coli AlaRS. Nine structures were solved with this construct. The kinetic parameters for amino-acid activation (adenylate synthesis) of the engineered and native protein were closely similar (Supplementary Figs 3 and 4). The structures reported here are generally referred to as AlaRS_{441-LZ} and G237A AlaRS_{441-LZ}, where G237A is a substitution that affects the shape of the amino-acid-binding pocket (see below). The structures of AlaRS_{441-LZ} with five different ligands were co-crystallized in two different crystal forms. Subsequently, G237A AlaRS_{441-LZ}, both alone and bound with three ligands, was crystallized (Supplementary Table 1). All complex structures were solved at about 2 Å resolution (1.9–2.2 Å) with an estimated atomic coordinate error of about 0.2 Å, with one exception (the complex of G237A AlaRS_{441-LZ} with the seryl-adenylate analogue, which was solved at 2.8 Å resolution).

E. coli AlaRS_{441-LZ} adopts a cradle-like shape consisting of the N-terminal (residues 1–249) seven-stranded β-sheet structure (with flanking α-helices) characteristic of class II tRNA synthetases, followed by the C-terminal α-helix bundle that forms a tRNA recognition motif (Fig. 1a). The three signature motifs (motifs 1, 2 and 3) of class II enzymes are located at residues 5–21, 60–97 and 233–248, respectively. The structure of the *E. coli* apoenzyme resembles that of *Aquifex aeolicus* AlaRS (root mean squared deviation (r.m.s.d.)

¹The Skaggs Institute for Chemical Biology and Department of Molecular Biology, The Scripps Research Institute, BCC-379, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. †Present address: Metabolon, Inc., 800 Capitola Drive, Suite 1, Durham, North Carolina 27713, USA.

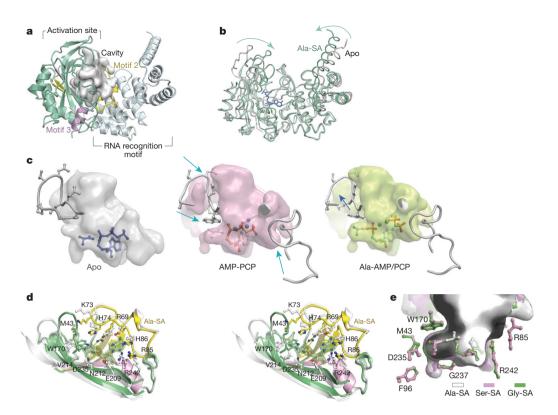


Figure 1 | Plasticity of active site in AlaRS. a, Overall structure of the N-terminal catalytic fragment AlaRS_{441-LZ} of *E. coli* AlaRS, shown with the water-accessible cavity (grey). b, Global movement of catalytic domain on ligand binding. c, Dynamic adjustments of the active-site pocket of AlaRS. Left, active-site cavity of apo-AlaRS_{441-LZ} with L-alanine and ATP placed by modelling (blue). The model of apo-AlaRS_{441-LZ} is derived from the structure of apo-G237A AlaRS_{441-LZ} by mutating Ala 237 back to Gly. Middle, active-site cavity of the AMP-PCP complex with bound AMP-PCP and L-alanine placed by modelling (grey). Right, active-site cavity of the Ala-AMP/PCP complex with bound Ala-AMP and βγ-PCP. The movements of

2.55 Å over 407 Cα positions). Three of the five complexes with wildtype AlaRS441-LZ included those with the ATP analogue adenyl 5'-(β,γ-methylene) triphosphate (AMP-PCP) alone, with the alanyl-AMP adenylate and the pyrophosphate (PP_i) analogue (β,γmethylene)pyrophosphate (βγ-PCP) all bound together, and with the adenylate analogue 5'-O-(N-(L-alanyl)-sulphamoyl adenosine) (Ala-SA). The structures of AlaRS_{441-LZ} in the complexes are closely similar to each other, with r.m.s.d. values of 0.35 Å between the two complexes with the adenylates, and a r.m.s.d. of 0.67 Å between the adenylate complexes and the ATP analogue complex. In contrast, much larger deviations (r.m.s.d. 2.6–2.8 Å over 418 Cα positions) were observed when comparing the complexes with the apoenzyme. In particular, the enzyme undergoes an apparent 'open to closed' transition on ligand binding. This contraction involved not only the activation site but also the C-terminal tRNA recognition motif (Fig. 1b).

The active-site cavity is mainly formed by motif 2, motif 3 and two loop regions connecting $\beta4$ – $\beta5$ and $\beta5$ – $\beta6$, respectively (Fig. 1a and Supplementary Fig. 3). The active-site cavities of the various enzyme structures were calculated for different states with the program Hollow¹⁹. The solvent-accessible surface around the active site was computed with a probe radius of 1.4 Å (approximating a water molecule). This shell defined the ligand-binding space formed by the enzyme. From this calculation, the cavity of the apoenzyme is seen clearly as wide open (Fig. 1c), with a space at the amino-acid-binding site much larger than is required to fit the cognate amino acid (alanine). Binding of AMP-PCP induced a marked contraction of the active centre that shrank the amino-acid-binding site and the adenylate-binding site (Fig. 1c). With this contraction, the space available for amino-acid

two active-site loops are indicated by arrows. **d**, Stereo view of 'open to closed' conformational change at the active site on Ala-SA binding. The difference map of the omitted Ala-SA ligand is contoured at 9.0σ (green). **e**, Amino-acid pocket with three different ligands bound in AlaRS_441-LZ. Three co-crystal structures of AlaRS_441-LZ with Ala-SA, Ser-SA and Gly-SA were superimposed. The ligand, active-site residues and cavity of each complex structure are coloured as indicated. The small cavity containing three water molecules is seen below the amino-acid moiety of all three bound ligands.

binding became even smaller than that needed to fit alanine. In particular, when bound to AMP-PCP, the space for the amino group of alanine was partly occupied by the indole side chain of Trp 170 (Fig. 1c). When alanine was bound to the amino-acid site together with ATP to form the Ala-AMP adenylate, Trp 170 swung back slightly (Fig. 1c, d). Thus, the enzyme did not preform a rigid amino-acid-binding pocket, but rather was adjustable to the binding ligands. Replacing Ala-AMP with the Ala-SA analogue resulted in a similar structure (r.m.s.d. 0.15 Å, for residues within 5 Å of the ligands).

Structures with Ser-SA and Gly-SA had similar resolutions (about $1.9\,\text{Å}$) and the same space group ($P4_12_12$) as the complex with Ala-SA. The structures of the three adenylate analogue complexes closely resemble each other (maximum r.m.s.d. 0.18 Å). The calculated active-site cavities showed no apparent differences between these three structures, and all cavity-forming residues aligned to the same positions (Fig. 1e). After superimposition of the three proteins, the adenylate parts of all three ligands—adenosine, ribose and sulphur also aligned closely. Thus, all three amino acids bound to almost identical positions in the active site (r.m.s.d. 0.14 Å). Only Cβ of serine rotated by 0.53 Å in comparison with Cβ of alanine. This rotation allowed the extra hydroxyl group of serine to fit snugly into the lower side of the amino-acid pocket. A small cavity right under the amino-acid site is partly filled by three stably bound water molecules (found in all three complex structures; Fig. 1e). No apparent change of pocket size occurred on binding of serine (no expansion) or glycine (no shrinkage).

Similarly to alanine, the α -amino group of serine forms a hydrogen bond with the carboxyl group of Asp 235 (Fig. 2a, b, d). In addition, the carboxyl group of serine forms a hydrogen bond with the class II

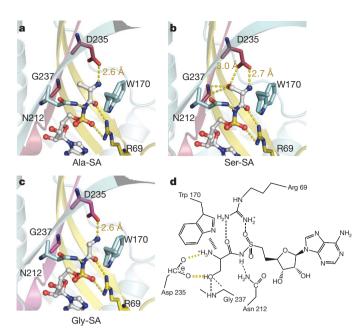


Figure 2 | Alanine active site with intrinsic design defect that mis-binds serine. a–c, Interactions with L-alanine, L-serine and glycine as shown in their ligand-bound structures. Two extra interactions are formed between AlaRS and the γ-OH of serine: Asp $235(COO^-)$ -Ser(OH) and Gly 237(N)-Ser(OH). A contact between Gly $237(C\alpha)$ -Ser(OH) is also shown (3.3 Å). Asp 235 also forms a common hydrogen bond with the α-amino group of all three amino acids. d, Diagram of the interactions between the amino-acid moiety of Ser-SA and the synthetase. The aromatic ring of Trp 170 stacks with the backbone of amino-acid substrates.

invariant Arg 69 on motif 2. The same Arg 69 forms a salt bridge with the α -sulphate of the adenylate analogue, and this interaction is critical for catalysing the activation reaction of class II enzymes^{20,21}. A conserved residue, Asn 212, is hydrogen-bonded with the bridging nitrogen between the seryl moiety and AMP (replaced by O in the

natural adenylate). When an amino acid was absent, the side chain of Asn 212 swung to bind to the α -phosphate of AMP-PCP. These interactions are also observed in the structures of AlaRS_{441-LZ} with Ala-SA and Gly-SA, with almost all coordinates of the AlaRS_{441-LZ} residues within the binding pocket being the same, regardless of the bound amino acid (Fig. 2a, c). This common main-chain conformation may explain why cognate and non-cognate amino acids have similar k_{cat} values and are differentiated mainly by their respective K_{m} values²².

In contrast with serine, α -aminobutyrate is not activated by AlaRS, possibly because the larger size of CH₃ (compared with that of OH) blocks the entry of α-aminobutyrate into the active site¹³. However, our structures show that serine forms interactions (with the enzyme) that are not possible with α-aminobutyrate. Two extra hydrogen bonds pin down the hydroxyl group. One is from the G237 backbone NH, whereas the other is from the carboxyl group of Asp 235 (Fig. 2b). As mentioned above, the carboxyl group of Asp 235 (conserved in all AlaRSs) also holds the α -NH₃⁺ group of serine, alanine and glycine. (This finding, in the present work, sheds light on the lethality of a previously described D235A substitution in E. coli AlaRS, which a kinetic analysis suggested was due in large part to the role of D235 in transferring the aminoacyl group from the adenylate to tRNA²³, probably by stabilizing the transition state²⁴.) The necessity of holding the α -amino group (of the cognate substrate alanine) through an acidic group therefore creates the dilemma of not being able to avoid serendipitous binding to the serine OH. The interactions flowing from Asp 235 are major determinants in the binding of serine to AlaRS (see below).

A close inspection of the structure showed that the serine OH is only 3.3 Å away from the $C\alpha$ of G237 in motif 3, β 7 (Fig. 2b). Binding of serine does not induce any apparent movement of β 7, which is the central β -strand in the class II-defining β -sheet cradle and is therefore the least flexible compared with other parts of the cradle. This lack of flexibility may explain why the $K_{\rm m}$ for serine is 600-fold that for alanine (Fig. 3a). Because of the close approach of the serine OH to the $C\alpha$ of G237, the pocket would probably be shrunk by substituting a slightly bulkier residue for G237. Pursuant to this idea, G237A AlaRS₄₄₁ (without a leucine zipper) was made.

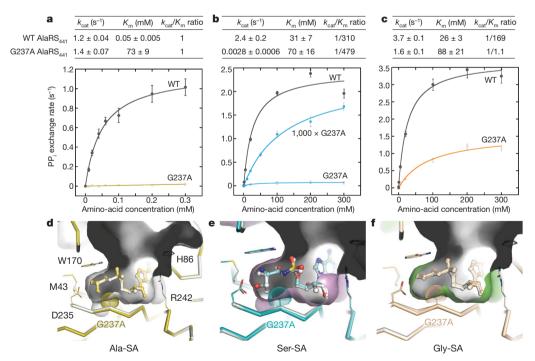


Figure 3 | Mutant enzyme with a smaller pocket retains serine binding affinity. **a–c**, Kinetic parameters for amino-acid activation of G237A and wild-type (WT) AlaRS₄₄₁. **a**, Alanine; **b**, serine; **c**, glycine. $k_{\text{cat}}/K_{\text{m}}$ ratios of non-cognate serine and glycine were calculated against cognate alanine for each protein, separately. Errors are s.d. $(n \ge 3)$. **d–f**, Comparison of G237A

AlaRS_{441-LZ} structures in complex with ligands Ala-SA (\mathbf{d}), Ser-SA (\mathbf{e}) and Gly-SA (\mathbf{f}) with the corresponding WT AlaRS_{441-LZ} complex structures (white). For clarity, only the active-site cavities of WT AlaRS_{441-LZ} structures are shown here.

In the standard amino-acid-dependent ATP–PP_i exchange assay (Fig. 3a–c), which follows the rate of synthesis of the adenylate, G237A AlaRS₄₄₁ had a sharply elevated $K_{\rm m}$ for alanine (1,250-fold increase) (Fig. 3a). This sharp drop in affinity for alanine was expected, because the distance between the C α of G237 and C β of alanine is 3.8 Å, and the extra β -CH₃ present in G237A AlaRS₄₄₁ would crowd the bound alanine. Next we found that G237A AlaRS₄₄₁ showed only a small change in the $K_{\rm m}$ for glycine. This result is consistent with the introduced extra β -CH₃ of G237A AlaRS₄₄₁ affecting only the β -position of a bound amino-acid substrate (Fig. 3c). Although serine is the bulkiest of the three amino acids, G237A AlaRS₄₄₁ bound to serine with a similar affinity as that of wild-type AlaRS₄₄₁ (Fig. 3b). Therefore, within limits, the pocket size does not have a key function in serine binding.

To achieve further clarity on the reason for serine being insensitive to the shrinkage of the pocket size, we crystallized G237A AlaRS_{441-LZ} bound to each of the analogues Ala-SA, Gly-SA and Ser-SA (Fig. 3d–f). As expected, the G237A mutation introduced bulk into the alanine-binding pocket, with little other change in the pocket or the surrounding atoms. The steric crowding of alanine with the introduced C β in G237A AlaRS_{441-LZ} is apparent in the small separation of C β of alanine from C β of A237 (3.6 Å; Fig. 3d). This crowding easily explains the sharply elevated $K_{\rm m}$ for alanine of G237A AlaRS_{441-LZ}.

As expected, bound Gly-SA was little affected by the extra bulk from the $\beta\text{-CH}_3$ introduced at position 237 in G237A AlaRS_{441-LZ} (Fig. 3f). For that reason, the $K_{\rm m}$ for glycine was virtually unchanged. However, for serine, crowding by the introduced C β of G237A AlaRS_{441-LZ} is clear. And yet, whereas crowding at the C β positions of A237 and serine occurs (which breaks the hydrogen bond between backbone NH of G237A and the serine OH and extends the separation to 4.0 Å), the carboxyl of Asp 235 remains hydrogen-bonded to the serine OH and the $\alpha\text{-NH}_3^+$ (Fig. 3e). Thus, the bifurcated interactions between Asp 235 and serine seem to be a major cause of the inability of the enzyme to evolve to a higher discrimination against serine.

Inspection of the k_{cat} and $k_{\text{cat}}/K_{\text{m}}$ values for the three amino acids, with native AlaRS₄₄₁ and G237A AlaRS₄₄₁, further showed the difficulty of resolving the dilemma posed by the structural design of the enzyme. Because of the extra bulk introduced by the G237A substitution, the carboxyl of serine shifts 0.3 Å towards motif 3 (Fig. 3e) and this shift may in turn explain why the k_{cat} for serine activation is greatly decreased (Fig. 3b). However, the desirable outcome of decreasing the kinetic

activation of serine comes at a high price. The shrinking of the pocket (in the hope of excluding serine) sharply raised the $K_{\rm m}$ for alanine but at the same time did not greatly perturb the $K_{\rm m}$ for serine, or did not provide a sufficient offsetting decrease in $k_{\rm cat}$ for serine. The problem is further compounded for glycine (as expected, the $k_{\rm cat}/K_{\rm m}$ ratio for glycine was least affected by the mutation), which is more than 150-fold more competitive against alanine with G237A AlaRS₄₄₁ than with wild-type AlaRS₄₄₁.

Because D235 is the major reason for the binding of serine in the pocket for alanine, we investigated whether this specificity problem could be solved by changing (D235E) or removing (D235N, D235Q) the acidic group or the side chain (D235A) of D235 (Fig. 4a, Supplementary Fig. 6 and Supplementary Information). None of these substitutions improved the discrimination between Ala and Ser. For example, the frequency of misincorporating serine, relative to incorporating alanine, was much greater with the D235N and D235A mutants than with the wild-type enzyme (Supplementary Figs 6 and 7). Collectively, these results further highlight the dilemma that creates this 'serine paradox', and how difficult it is to bypass Asp 235.

While they are continuously challenged by canonical and non-canonical amino acids, AARSs control the fidelity of the genetic code on the protein synthesis production line. Only AlaRS and ThrRS are known to misactivate both smaller and larger non-cognate amino acids. However, the problematic larger amino acid (hydroxynorvaline) for ThrRS does not occur naturally²⁵. For that reason, ThrRS has no selective pressure for rejecting hydroxynorvaline. In contrast, AlaRSs are challenged by canonical amino acids, both the smaller glycine and the larger serine.

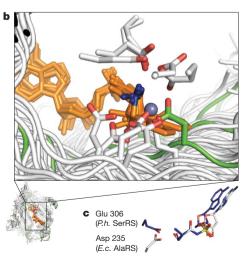
The work described above shows that this serine paradox is rooted in a dilemma that flows out of the ancient design of AlaRSs. For both class I and class II enzymes, the ancient scaffold simultaneously captures the α -amino and α -carboxyl groups of the amino-acid substrate. A natural consequence of this circumstance is the need for an acidic residue to pin down the α -NH₃⁺ group. In contrast with previous understanding²⁶, the way in which this is done is idiosyncratic to the synthetase, and many synthetases have found alternative residues during evolution (Fig. 4b and Supplementary Fig. 8). Asp 235 is universally conserved in AlaRSs across all three kingdoms of life. This deeply rooted architecture causes the serine-interaction paradox for AlaRSs, and the interaction of Asp 235 with the serine OH is so robust that it was retained even after the creation of a shrunken pocket with the G237A substitution (Fig. 3). Similarly to

a			
Alanine	k _{cat} (s ⁻¹)	K _m (mM)	k _{cat} /K _m ratio
WT AlaRS ₄₄₁	1.2 ± 0.04	0.05 ± 0.005	1
D235A AlaRS ₄₄₁	0.069 ± 0.01	502 ± 115	1
D235N AlaRS ₄₄₁	0.074 ± 0.01	178 ± 14	1

Serine	$k_{\rm cat}$ (s ⁻¹)	$K_{\rm m}$ (mM)	$k_{\rm cat}/K_{\rm m}$ ratio
WT AlaRS ₄₄₁	2.4 ± 0.2	31 ± 7	1/310
D235A AlaRS ₄₄₁	n.d.	>600	-
D235N AlaRS ₄₄₁	0.0046 ± 0.002	58 ± 9	1/5.2

Glycine	k _{cat} (s ⁻¹)	K _m (mM)	k _{cat} /K _m ratio
WT AlaRS ₄₄₁	3.7 ± 0.1	26 ± 3	1/169
D235A AlaRS ₄₄₁	0.0035 ± 0.001	493 ± 82	1/19
D235N AlaRS ₄₄₁	0.0058 ± 0.002	42 ± 4	1/3.0

Figure 4 | **Critical role of Asp 235 shows dilemma for AlaRSs. a**, Kinetic parameters for amino acid activation of D235A and D235N. n.d., not determined. Errors are s.d. ($n \ge 3$). **b**, Various designs of acidic residues that recognize the α-amino group of amino acids in class II AARSs; the structures are aligned by their amino acid substrates (orange, with α-amino groups in



blue). The central β -sheets are shown (AlaRS in green; others in white). The acidic residues are shown with carboxyl groups in red. **c**, Ser-SA ligands in the active site of *E. coli (E.c.)* AlaRS and *Pyrococcus horikoshii (P.h.)* SerRS²⁹. Both acidic residues form bifurcated interactions with the α -amino group and γ -OH of serine.

AlaRS, the groups that bind the α -NH₃⁺ group in SerRS and ThrRS (glutamate or zinc) are also used to recognize the γ -OH group of their cognate substrates^{25,27} (Fig. 4c).

The structural architecture of class II (and class I) tRNA synthetases goes back to the last universal common ancestor about 3.5 billion years ago. Asp 235 has been in place in AlaRSs for that entire time. Despite billions of years of evolution, an alternative to Asp 235, within the context of the structural scaffold of class II enzymes, has not been found. Apparently unable to find a better architecture for the recognition of alanine, the problem of serine misactivation was solved through the free-standing AlaXps. (Indeed, in a bacterial model, AlaXp rescued a cell from serine toxicity¹⁵.) These proteins appeared contemporaneously with the early AlaRSs¹⁷. No doubt this solution removed much of the evolutionary pressure to find a variation of the structural architecture of AlaRS that would solve the serine paradox (Supplementary Fig. 1).

METHODS SUMMARY

To aid crystallization, a 'leucine-zipper' mutation²⁸ (His104Leu, Gln108Leu, Glu112Leu) was created in an active 441-residue fragment of E. coli AlaRS. Wild-type and mutant E. coli AlaRS_{441-LZ} proteins were overexpressed from E. coli and purified by fast protein liquid chromatography. The ligand complexes were generated by mixing at 4 °C the protein with Ala-SA, Ser-SA, Gly-SA, AMP-PCP, or alanine and AMP-PCP together, with 10 mM MgCl₂ and 10 mM 2-mercaptoethanol. Crystals were obtained by the micro-sitting-drop vapour diffusion method. All wild-type and mutant AlaRS_{441-LZ} complex structures were determined by molecular replacement with the apo-AlaRS structure from A. aeolicus (Protein Data Bank accession code 1RIQ) as the initial model. The ATP-pyrophosphate exchange assays were performed at 22 °C in a buffer composed of 100 mM Tris-HCl pH 8.0, 1 mM ATP, 10 mM MgCl₂, 10 mM KF, 0.1 mg ml⁻¹ BSA, 5 mM 2-mercaptoethanol and 0.5 mM Na-³²PP_i. Linear initial rates were determined by measurements of the incorporation of $^{\dot{32}}PP_{i}$ into ATP at four sequential time intervals of 3 min each, and the data were fitted to the Michaelis-Menten equation with KaleidaGraph.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 July; accepted 26 October 2009.

- Beebe, K., Ribas De Pouplana, L. & Schimmel, P. Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. EMBO J. 22, 668–675 (2003).
- Lee, J. W. et al. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature 443, 50–55 (2006).
- Beebe, K., Mock, M., Merriman, E. & Schimmel, P. Distinct domains of tRNA synthetase recognize the same base pair. *Nature* 451, 90–93 (2008).
- Carter, C. W. Jr. Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. *Annu. Rev. Biochem.* 62, 715–748 (1993).
- Giege, R. The early history of tRNA recognition by aminoacyl-tRNA synthetases. J. Biosci. 31, 477–488 (2006).
- Norris, A. T. & Berg, P. Mechanism of aminoacyl RNA synthesis: studies with isolated aminoacyl adenylate complexes of isoleucyl RNA synthesise. *Proc. Natl Acad. Sci. USA* 52, 330–337 (1964).
- Eldred, E. W. & Schimmel, P. R. Rapid deacylation by isoleucyl transfer ribonucleic acid synthetase of isoleucine-specific transfer ribonucleic acid aminoacylated with valine. J. Biol. Chem. 247, 2961–2964 (1972).
- Boniecki, M. T., Vu, M. T., Betha, A. K. & Martinis, S. A. CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase. *Proc. Natl Acad.* Sci. USA 105, 19223–19228 (2008).
- 9. Fersht, A. R. Enzyme Structure and Mechanism (Freeman, 1977).
- Nureki, O. et al. Structural basis for amino acid and tRNA recognition by class I aminoacyl-tRNA synthetases. Cold Spring Harb. Symp. Quant. Biol. 66, 167–173 (2001).

- 11. Fersht, A. R. Sieves in sequence. Science 280, 541 (1998).
- Fukai, S. et al. Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA^{Val} and valyl-tRNA synthetase. Cell 103, 793–803 (2000).
- Tsui, W. C. & Fersht, A. R. Probing the principles of amino acid selection using the alanyl-tRNA synthetase from *Escherichia coli. Nucleic Acids Res.* 9, 4627–4637 (1981).
- Ahel, I., Korencic, D., Ibba, M. & Söll, D. Trans-editing of mischarged tRNAs. Proc. Natl Acad. Sci. USA 100, 15422–15427 (2003).
- Chong, Y. E., Yang, X. L. & Schimmel, P. Natural homolog of tRNA synthetase editing domain rescues conditional lethality caused by mistranslation. *J. Biol. Chem.* 283, 30073–30078 (2008).
- 16. Ling, J. et al. Resampling and editing of mischarged tRNA prior to translation elongation. Mol. Cell 33, 654–660 (2009).
- Guo, M. et al. The C-Ala domain brings together editing and aminoacylation functions on one tRNA. Science 325, 744–747 (2009).
- Sokabe, M., Okada, A., Yao, M., Nakashima, T. & Tanaka, I. Molecular basis of alanine discrimination in editing site. *Proc. Natl Acad. Sci. USA* 102, 11669–11674 (2005)
- 19. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
- Arnez, J. G. & Moras, D. Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.* 22, 211–216 (1997).
- Davis, M. W., Buechter, D. D. & Schimmel, P. Functional dissection of a predicted class-defining motif in a class II tRNA synthetase of unknown structure. *Biochemistry* 33, 9904–9911 (1994).
- Jakubowski, H. in The Aminoacyl-tRNA Synthetases (eds Ibba, M., Francklyn, C. & Cusack, S.) 384–396 (Eurekah, 2005).
- 23. Shi, J. P., Musier-Forsyth, K. & Schimmel, P. Region of a conserved sequence motif in a class II tRNA synthetase needed for transfer of an activated amino acid to an RNA substrate. *Biochemistry* **33**, 5312–5318 (1994).
- Xin, Y., Li, W. & First, E. A. Stabilization of the transition state for the transfer of tyrosine to tRNA^{Tyr} by tyrosyl-tRNA synthetase. *J. Mol. Biol.* 303, 299–310 (2000).
- Sankaranarayanan, R. et al. Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. Nature Struct. Biol. 7, 461–465 (2000).
- First, E. A. in *The Aminoacyl-tRNA Synthetases* (eds Ibba, M., Francklyn, C. & Cusack, S.) 328–352 (Eurekah, 2005).
- Belrhali, H. et al. Crystal structures at 2.5 angstrom resolution of seryl-tRNA synthetase complexed with two analogs of seryl adenylate. Science 263, 1432–1436 (1994).
- 28. Cieslik, M. & Derewenda, Z. S. The role of entropy and polarity in intermolecular contacts in protein crystals. *Acta Crystallogr. D* **65**, 500–509 (2009).
- Itoh, Y. et al. Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon Pyrococcus horikoshii. RNA Biol. 5, 169–177 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. J. Read for pointing out the errors in the previous AlaRS-ligand structures and for crystallographic discussions; and G. J. Kleywegt, Z. Otwinowski and A. Perrakis for technical assistance. X-ray diffraction data were collected at Stanford Synchrotron Radiation Laboratory (SSRL) beamlines 7-1, 9-1 and 11-1. This work was supported by grant GM 15539 from the National Institutes of Health and by a fellowship from the National Foundation for Cancer Research.

Author Contributions M.G., X.-L.Y. and P.S. designed the experiments. M.G., Y.E.C., R.S. and K.B. performed the experiments. M.G. and Y.E.C. analysed the data. M.G., Y.E.C., X.-L.Y. and P.S. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3hxu (WT/Ala-SA), 3hxv (WT/Gly-SA), 3hxw (WT/Ser-SA), 3hxx (WT/AMP-PCP/Mg(II)), 3hxy (WT/Ala-AMP/PCP/AMP-PCP/Mg(II)), 3hxz (G237A/Ala-SA), 3hy0 (G237A/Gly-SA) and 3hy1 (G237A-apo and G237A/Ser-SA). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.S. (schimmel@scripps.edu).

doi:10.1038/nature08612 nature

METHODS

Preparation of proteins for crystallization. A leucine-zipper mutant was introduced on the N-terminal aminoacylation domain of E. coli AlaRS, on the basis of rational mutagenesis designed to aid crystallization of the E. coli AlaRS protein (details available from the authors on request). Plasmids for expression of E. coli AlaRS fragments were constructed through PCR amplification of the desired region using primers containing NdeI-XhoI sites and ligated into pET20b to generate pET20b-EcAlaRS₀₋₄₄₁, which encodes the AlaRS fragment without a tag (as with endogenous E. coli AlaRS, the initial methionine, Met 0, is cleaved in E. coli during expression, and the mature sequence starts from the following serine, Ser 1). The leucine-zipper mutant was created by mutating residues His 104, Gln 108 and Glu 112 into leucines. This AlaRS_{441-LZ} protein was expressed in E. coli at 23 °C by induction with 0.1 mM isopropyl β-D-thiogalactoside, and purified sequentially on a DEAE column, a Q high-performance column, and a phenol high-performance column (GE Healthcare). Purified protein was dialysed against 5 mM Tris-HCl buffer pH 8.0 containing 50 mM NaCl and 1 mM 2-mercaptoethanol. The G237A mutant AlaRS_{441-LZ} was purified similarly.

Crystallization, data collection, structure determination and refinement. A previous and limited analysis of a co-crystal of Aquifex aeolicus AlaRS with alanine resolved the interactions of the cognate amino acid with the protein at a resolution of 2.08 Å (ref. 30). Although the structure itself was accurate, it later came to our attention that the details of the proposed interactions with free serine were incorrect (see newly deposited Protein Data Bank coordinates under accession code 3htz). In addition, we were concerned about the observation that, because of their small size and the absence of the AMP moiety, co-crystallized amino acids can bind to alternative sites within the active centre³¹. We therefore set out to obtain co-crystals of aminoacyl adenylates with E. coli AlaRS. Purified ${\rm AlaRS_{441\text{-}LZ}}$ was then mixed with different ligands $(40\,{\rm mg\,ml}^{-1}$ protein, with 2 mM Ala-SA/Ser-SA/Gly-SA/AMP-PCP, or with 20 mM L-alanine and 2 mM AMP-PCP), together with 10 mM MgCl2 and 10 mM 2-mercaptoethanol, at 4 °C. Crystallization was performed by the micro-sitting drop method. A drop was prepared by mixing 0.1 µl of protein solution with 0.1 µl of precipitant solution and was equilibrated against 70 µl of precipitant solution. Ala-SA, Ser-SA and Gly-SA complex crystals were crystallized at 4 °C under 27-32% PEG400, 100 mM HEPES-Na pH 7.5-8.0. AMP-PCP crystals were crystallized at 4 °C under 20% PEG400, 0.2 M sodium sulphate. Ala/AMP-PCP crystals were crystallized at 4 °C under 20% PEG3000, 0.1 M Tris-HCl pH 8.5. Crystals were collected after incubation for 5-14 days and were flash-frozen at 100 K for data collection.

The data sets were obtained from SSRL beamlines 11-1 and 9-1 and were processed with HKL2000 (ref. 32). The initial Ala-SA complex structure was solved by molecular replacement with the $A.\ aeolicus$ AlaRS structure (PDB accession code Iriq). Iterative model building and refinement was performed with Coot and Refmac5 (refs 33, 34), with the Ala-SA ligand added only after the R-factor was lower than 25%. The current model has $R_{\rm work}$ and $R_{\rm free}$ values of 15.3% and 19.0%, respectively. Inspection of this structure indicated that both the introduced leucine-zipper and the end residue (Phe 441) formed an extensive interface for crystal packing and were crucial for crystallization. AlaRS $_{441-LZ}$ structures of other ligand complexes were solved by using the Ala-SA complex as the initial model. Each structure has only one AlaRS $_{441-LZ}$ molecule per asymmetric unit (Supplementary Table 1). Inspection of the difference density of the Ala/AMP-PCP complex indicated that these substrates partly reacted in the

crystal, with clear density of dissociating $\beta\gamma$ -PCP. The current model has an AMP-PCP, an Ala-AMP and a $\beta\gamma$ -PCP within the same active site, each with occupancy of 0.5.

G237A AlaRS $_{441\text{-LZ}}$ in complex with Ala-SA or Gly-SA was crystallized under

similar conditions to those of wild-type AlaRS_{441-LZ} by mixing with 2 mM Ala-SA

or Gly-SA. G237A AlaRS_{441-LZ} in complex with Ser-SA could only be crystallized by mixing with 10 mM Ser-SA, under 25% PEG550 MME, 0.1 M MES pH 6.5, at room temperature. Data on the crystals of the G237A AlaRS_{441-LZ} complexes was collected at SSRL beamline 7-1. In contrast with the wild-type AlaRS_{441-LZ} crystals, each ligand induced the G237A mutant crystal into a distinct space group, with two molecules per asymmetric unit for Gly-SA (P2₁2₁2₁), and four for Ala-SA (C2). The molecules within each asymmetric unit closely resembled each other by non-crystallographic symmetry. The Ser-SA G237A mutant crystal contained two different molecules in one asymmetric unit, including one in complex with Ser-SA and another in the apo form. This difference indicates that the G237A mutant has a lower binding affinity for the Ser-SA analogue, which is consistent with our observation that the G237A mutation caused a sharp decrease in the rate of Ser-AMP formation. All ligands were added only after the R-factor was lower than 25% (28% for the Ser-SA G237A structure, 2.8 Å). All current models have good geometry and no residues in disallowed regions of the Ramachandran plot. Data collection and refinement statistics are given in Supplementary Table 1. Preparation of proteins for kinetic assays. The E. coli AlaRS fragment of interest was isolated through PCR amplification and inserted into pET20b to generate pET20b-His-EcAlaRS $_{0-441}$, with a $6 \times$ His tag at the N terminus of the AlaRS fragment, but without the leucine-zipper mutations. The G237A, D235A, D235N, D235Q, D2325E and D235AG237A mutations were introduced by the Quikchange method (Stratagene) and verified by DNA sequencing. The D235E mutant was insoluble. All other proteins were expressed and purified by a gravity column packed with Ni²⁺-nitrilotriacetate beads (Qiagen), washed, and eluted with 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM imidazole. Proteins were then dialysed against 5 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM 2-mercaptoethanol. ATP-pyrophosphate exchange assays. The ATP-pyrophosphate exchange assays were performed at 22 °C in a buffer composed of 100 mM Tris-HCl pH 8.0, 1 mM ATP, 10 mM MgCl₂, 10 mM KF, 0.1 mg ml $^{-1}$ BSA, 5 mM 2-mercaptoethanol and 0.5 mM Na- 32 PP $_{\rm i}$. Exchange assays contained 100 nM enzyme for the alanine, glycine and serine substrates for both native and G237A mutant, or up to $148\,\mu\text{M}$ G237A mutant for serine substrates. Initial rates were determined over 3-min linear intervals using four points and fitted to the Michaelis-Menten equation with KaleidaGraph. The values were determined

 Swairjo, M. A. & Schimmel, P. R. Breaking sieve for steric exclusion of a noncognate amino acid from active site of a tRNA synthetase. *Proc. Natl Acad. Sci.* USA 102, 988–993 (2005).

from three independent experiments.

- 31. Malde, A. K. & Mark, A. E. Binding and enantiomeric selectivity of threonyl-tRNA synthetase. *J. Am. Chem. Soc.* **131**, 3848–3849 (2009).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
- CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763 (1994).
- 34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).



The San Francisco Bay area has long been a hub for cutting-edge technologies. Now it is one of the nation's leaders in clean-energy research. **Corinna Wu** reports.

he competition for US\$500 million of bioenergy research funding from oil giant BP was fierce and international. BP's aim was to find an institution that could manage a major bioenergy operation and readily collaborate with industry — all in a location that would attract the necessary expertise. In the end, a partnership between the University of California, Berkeley; Lawrence Berkeley National Laboratory in California; and the University of Illinois at Urbana-Champaign won the 2006 competition, beating the University of California, San Diego; Massachusetts Institute of Technology in Cambridge; Imperial College London; and the University of Cambridge, UK. That victory would provide the San Francisco Bay area, already known for its biotechnology innovations and computer chips, with a significant boost to its growing reputation as an energy-research hub.

The San Francisco Bay region has long been a haven for high-tech giants such as Google, innovators such as Apple and biotech behemoths such as Genentech. Now its entrepreneurial spirit has turned towards technologies to address global climate change, particularly in the realm of renewable energy. Research into clean energy is thriving, and there are opportunities, despite California's sagging economy.

The BP support has been used to create the Energy Biosciences Institute (EBI) in Berkeley,

which focuses on applying biological solutions to energy-sector needs. The institute is trying to develop lignocellulosic ethanol from *Miscanthus* and other perennial grasses. These plants are grown in Illinois, then shipped to Berkeley, where they are milled into different particle sizes and used as raw material for laboratory experiments. EBI scientists are

studying the best way to break down the plant matter into sugars that can then be fermented into ethanol.

Change of tack

Charlie Anderson was one of the first postdoctoral scholars at the EBI when it opened its doors two years ago. Anderson earned his doctorate at Stanford University in Palo Alto, California, in cellular and molecular biology, focusing on mammalian systems. As he neared graduation, he met the EBI's director Chris

Somerville, who was then at Stanford serving as director of the Carnegie Institution for Science's department of plant biology.

"I saw the possibility for climate change causing major disruptions in populations, especially poor populations," says Anderson, who had previously intended to do biomedical research. Recognizing that climate change could cause human suffering on a grand scale, he decided that he could help more effectively by working out how to reduce greenhouse-gas emissions.

The EBI is currently housed in two locations on the University of California, Berkeley campus, occupying one floor of the chemistry building as well as a cylindrical structure that used to be part of the Lawrence Berkeley

National Laboratory. There are plans to move it to a new 6,000-square-metre facility — complete with open labs and ample common space — on the northwest corner of the campus in 2013. The facility will have space for 120 full-time researchers, including up to 50 scientists from BP.

The EBI scientists' backgrounds span a wide range of disciplines, including plant biology, biochemistry, agricultural science, chemical engineering, public health, law and economics. The last

group is taking on projects such as how to model carbon emissions, land use and crop production, and how biofuels development could affect energy and food prices.

Like the EBI, the US Department of Energy's Joint BioEnergy Institute (JBEI) in nearby Emeryville works on biofuels. It occupies the fourth floor of a gleaming office building in an area full of old factories and



Charlie Anderson: working to reduce emissions.

CAREERS

warehouses, close to drug companies such as Novartis and Bayer. Whereas the EBI's focus is primarily on ethanol, the JBEI is also looking at turning plant matter into biogasoline — hydrocarbons that can be put straight into a vehicle's petrol tank — by engineering microbes. The institute has 170 researchers and technical staff with backgrounds similar to those of EBI researchers.

Such efforts have made the Bay area a leader in clean technology — knowledge-based work aimed at reducing energy use, waste and environmental impact. According to research firm Clean Edge in Portland, Oregon, government, academic and private-sector opportunities have combined to make the San Francisco–San Jose–Oakland region the top metropolitan area in the United States for clean-tech jobs, both in research and non-research activities (see 'Selling sunshine').

Companies find the area attractive because of the large pool of technical talent, the number of venture capitalists based there and its strong culture of innovation. Its reputation as a pleasant place to live is also a big draw, notes Tai Bethune, senior recruiter at solarcell manufacturer SunPower in San Jose.

Investments in the area are sizeable. Stanford University announced in January that it was establishing the Precourt Institute for Energy, a \$100-million centre as an umbrella organization for all its energyrelated research. The money will help fund 8-10 new faculty positions, 5 postdoctoral fellowships and about 20 graduate fellowships, says Franklin Orr, director of the new institute. Areas of focus include photovoltaics, biofuels, geothermal energy, advanced batteries, energy efficiency and improving the electric power grid. Stanford, meanwhile, spends another \$30 million on energy research in all these areas each year. Along with the newly created TomKat Center for Sustainable Energy, two existing Stanford programmes will be rolled into the



Big plans: Chris Somerville settles in at the EBI.

new centre: the Precourt Energy Efficiency Center and the Global Climate & Energy Partnership (GCEP).

Since 2002, the GCEP has given more than \$85 million of support to several dozen projects at Stanford and 23 partner institutions, cutting across many academic departments including electrical engineering, materials science, chemistry, physics and the geosciences, says managing director Richard Sassoon. Its goal is to do fundamental research leading to technologies that will reduce greenhouse-gas emissions.

Like the EBI, the partnership has explicit connections with industry. Four companies — ExxonMobil, General Electric, Schlumberger and Toyota — have pledged a total of \$225 million to the partnership. GCEP projects include research into improving the electrodes for solar cells, better hydrogen storage methods and technology aimed at sequestering the carbon emissions produced by coal-fired power plants.

"It's the type of research that industry itself probably wouldn't do, because it's too long term," Sassoon says. He adds that even if some projects don't succeed, the institute will help produce the next generation of energy researchers. The GCEP has more than 300 graduate students and postdoctoral scholars. Industry partners have hired students and postdocs from the GCEP, but there hasn't been any organized effort to link them with job opportunities, Sassoon says. It has happened more through informal networking.

Industry influence

Some are concerned about the explicit collaboration between industry and academia — in particular, the influence BP will wield in the case of the EBI (see Nature 445, 688-689; 2007). But industry involvement does have its employment advantages. Although most of California's higher-education and research programmes have been hit hard by a statewide budget crisis, the EBI has not been affected, according to its assistant director Susan Jenkins. BP remains committed to funding the institute at the levels announced at its opening. And as the EBI does not rely on state funding, no one paid through the EBI has been forced to take off as many as 26 unpaid days per year, as other University of California employees have had to do.

In many ways, research into clean energy is still in its infancy, and Anderson is interested to see where it all leads. "I don't know if it will be a situation in which large companies swallow up the small start-up companies and continue to do things in a corporate way, or if it's going to be some bioenergy equivalent of Google starting up and growing to be a huge conglomerate," he says. Given the rich history of high tech in the area, perhaps it will be both.

Corinna Wu is a freelance writer based in Oakland, California.

SELLING SUNSHINE

Libby Wayman earned a master's degree in mechanical engineering from Massachusetts Institute of Technology in Cambridge and came to California two years ago to work at the clean-tech firm SunPower in San Jose, California. This summer, she left to join a solar start-up called SunPrint in Richmond, located in the North Bay outside San Francisco.

Being in the renewable-energy field has allowed her to work on rapidly progressing technology and witness companies as they grow. "I wouldn't have got that if I had, say, been a mechanical engineer in the automotive industry," she says. Interest in photovoltaics and solar

power seems a natural outgrowth of Silicon Valley's semiconductor legacy. Instead of using their expertise to make smaller, faster chips, many electrical engineers and materials scientists have begun applying their training to making cheaper, more efficient photovoltaics.

Already, the Department of Energy's Lawrence Berkeley National Laboratory (LBNL) has invested heavily in renewable energy research, thanks to the efforts of former lab director and current energy secretary, Steven Chu. The Solar Energy Research Center is part of the LBNL's Helios vision to develop

fuels from sunlight. The centre concentrates on developing photoelectrochemical cells that perform artificial photosynthesis.

But solar power is already scaling up on a grand scale in California, with the San Francisco Bay area as a major hub.

Spurred by state incentives, Pacific Gas & Electric has contracted BrightSource Energy of Oakland, California, to build seven solar thermal-power plants. The output works out to 3,666 gigawatt-hours a year — enough to power 530,000 homes. The first power plant will be built in the Mojave Desert, and its first phase is expected to be completed in 2011. Once this first

plant is running, it will double the amount of solar thermal power being produced in the United States.

SunPower, meanwhile, designs and installs solar-power systems at all scales, from power plants to commercial buildings to private residences. With more than 5,000 employees, it is one of the largest clean-tech employers in the world. The company hires scientists and engineers at all levels of experience across a number of disciplines, including mechanical engineering, electrical engineering, physics and materials science.

"We've hired about 30 R&D people over the past 6 months," says senior recruiter Tai Bethune. **C.W.**

Evidence for escape from adaptive conflict?

Arising from: D. L. Des Marais & M. D. Rausher Nature 454, 762-765 (2008)

Gene duplication is the primary source of new genes¹, but the molecular evolutionary mechanisms underlying functional divergence of duplicate genes are not well understood². Des Marais and Rausher³ argued that data from plant dihydroflavonol-4-reductase (DFR) genes support the model that gene duplication allows the escape from adaptive conflict (EAC) among several functions of a single-copy progenitor gene⁴. As the authors indicated, the key predictions of EAC, in comparison to other models^{1,5,6}, are that (i) adaptive changes occur in both daughter genes after duplication, and (ii) these adaptive changes must improve ancestral functions. Furthermore, EAC indicates that (iii) the improvement of several ancestral functions is constrained before duplication, although this last point was not explicitly stated. Here we show that contrary to the predictions of EAC, only one of the duplicated DFR lineages exhibited adaptive sequence changes. Owing to the lack of information on enzyme concentrations³ we question the accuracy of enzyme activity comparisons, and it is thus not clear that any ancestral function has been improved in either lineage.

To test the predictions for one daughter lineage, the authors first investigated patterns of DNA sequence evolution³. It was shown that after the duplication of the progenitor *DFR* gene, the lineage leading to the common ancestors of *DFR-A* and *DFR-C* exhibited a non-synonymous-to-synonymous rate ratio significantly exceeding 1, indicative of adaptive evolution by positive selection. However, subsequent biochemical assays showed that these enzymes have "essentially no activity" on any ancestral substrate³ and therefore do not demonstrate improvement of ancestral function, violating the above criterion (ii).

The other daughter gene from the duplication, DFR-B, showed no evidence of positive selection from sequence analysis³, calling into question whether adaptive evolution has occurred. To test potential improvement of ancestral enzyme function after duplication, the authors made recombinant proteins from a heterologous system and measured catalytic activities on five substrates. Although the same volume of cell culture extract was used in all assays, the recombinant proteins were neither quantified nor equalized in concentration. It is known that the protein yield from heterologous systems can vary by several orders of magnitude, depending on a wide variety of factors⁷. In fact, an earlier study in a heterologous system nearly identical to the one used by Des Marais and Rausher reported very different yields for different DFR enzymes8. Thus, it is crucial to quantify the concentration of the obtained recombinant protein for any quantitative measure of activity, as has been done in previous studies of duplicate gene evolution^{9,10}. Without enzyme concentration information, the apparently higher activity of DFR-B depicted in their Fig. 2 could be due to a larger amount of enzyme rather than a truly higher enzyme activity as the authors interpreted³.

One may argue that the ratios of activities against different substrates can be compared among enzymes to control for differences in protein concentration. For example, one could infer from their Fig. 2 (ref. 3) that the relative activity on dihydrokaempferol (DHK) to dihydroquercetin (DHQ) is enhanced in DFR-B, compared to that of the progenitor enzyme. However, such relative activity information does not indicate whether the enhancement is due to increased activity on DHK, reduced activity on DHQ, or both. It could also be posited that the activity of DFR-B on DHQ is unlikely to be reduced compared to the progenitor enzyme if DHQ is the main precursor of most anthocyanins produced by post-duplication species. However, because Michaelis—Menten kinetic parameters were uncharacterized for these enzymes, even if the activity of DFR-B is lower than that of the progenitor, it remains possible that it is high enough for normal

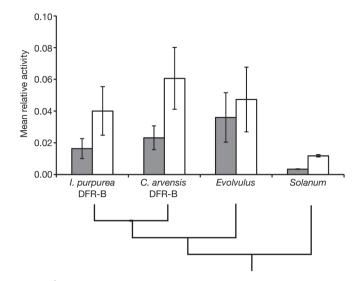


Figure 1 | Results of enzyme activities against different substrates. Shown are enzyme activities against naringenin (grey bars) and eriodyctiol (white bars), relative to those against DHK, DHQ and DHM. A greater number indicates a relative preference for narigenin (or eriodyctiol) compared to DHQ, DHK and DHM. Error bars denote one standard error. For a given enzyme, we calculate its activities against naringenin (or eriodyctiol) relative to DHK, DHQ and DHM, and show the average of the three relative activities and its standard error. The evolutionary relationships among the four enzymes are shown by the phylogeny. The relative activities are calculated from the data presented in Fig. 2 of ref. 3, graciously supplied by the authors.

physiology, particularly if DHQ concentrations are high. Furthermore, one cannot exclude the possibility that the activity on DHQ is retained in DFR-A and/or DFR-C but reduced in DFR-B. Even if different enzymes are compared using these ratios, it is clear that DFR-B is no better with naringenin and eriodyctiol (the less typical but proposed to be improved-upon substrates) compared to DHK, DHQ and dihydromyricetin (DHM) (the typical substrates of DFR) than the progenitor enzyme from *Evolvulus* (Fig. 1). Thus, there does not seem to be any unambiguous evidence for improvement of any ancestral function in DFR-B. Coupled with the failure to detect positive selection in DFR-B, these results go against both criteria (i) and (ii). Because there is no evidence of improvement of ancestral function in any of DFR-A, -B and -C, criterion (iii) obviously cannot be established.

In conclusion, none of the three key features of the EAC model have been clearly demonstrated in the evolution of DFRs. Although EAC may be a valid model for describing gene family evolution, rigorous tests of the predictions associated with it are needed to describe its general importance.

Todd Barkman¹ & Jianzhi Zhang²

¹Department of Biological Sciences, Western Michigan University, Kalamazoo, Michigan 49008, USA.

²Department of Ecology and Evolutionary Biology, University of Michigan, 1075 Natural Science Building, 830 North University Avenue, Ann Arbor, Michigan 48109, USA.

e-mail: jianzhi@umich.edu

Received 28 October 2008; accepted 14 August 2009.

- 1. Ohno, S. Evolution by Gene Duplication (Springer-Verlag, 1970).
- Zhang, J. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298 (2003).
- 3. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).

- 4. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **256**, 119–124 (1994).
- He, X. & Zhang, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169, 1157–1164 (2005)
- Force, A. et al. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545 (1999).
- Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 72, 211–222 (2006).
- Shimada, N. et al. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the Lotus japonicus genome. J. Exp. Bot. 56, 2573–2585 (2005).
- Thomson, J. M. et al. Resurrecting ancestral alcohol dehydrogenases from yeast. Nature Genet. 37, 630–635 (2005).
- Zhang, J. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nature Genet. 38, 819–823 (2006).

Competing financial interests: declared none

doi:10.1038/nature08663

Des Marais & Rausher reply

Replying to: T. Barkman & J. Zhang Nature 462, doi:10.1038/nature08663 (2009)

Barkman and Zhang¹ level two criticisms at our report of escape from adaptive conflict (EAC)². The first criticism is that our evidence of repeated adaptive substitution in the *DFR-A/C* lineage is consistent with either EAC or neofunctionalization in that lineage. As indicated in our original report², we do not disagree with this claim, but believe it misses the point that the hallmark of EAC is the adaptive improvement of function in both duplicate lineages. The repeated positive selection in the *DFR-A/C* lineage simply demonstrates improvement of some as yet unspecified function in that lineage.

Their second criticism¹ is that we failed to quantify enzyme concentration in our assays, and because of this our results do not demonstrate improvement of function in the *DFR-B* lineage. However, an examination of the possible effects of uncertainty in enzyme concentration indicates that our inference of improved catalytic function on anthocyanin precursors at the base of the *DFR-B* lineage remains robust. We can represent uncertainty in enzyme concentration by the relationship

$$C_1 = \gamma C_{\rm B} \tag{1}$$

in which C_1 and C_B are the absolute enzyme concentrations in the assays using the single-copy DFR and the DFR-B copy, respectively. Furthermore, we represent the specific activity of enzyme type i (single-copy or B-copy) on substrate k by A_i^k , in which specific activity is defined by

$$P_i^k = C_i A_i^k \tag{2}$$

Here P_i^k is the concentration of the anthocyanidin reaction product produced by enzyme i acting on substrate k. Combining equations (1) and (2) yields

$$A_{\rm R}^k = \gamma A_1^k \left(P_{\rm R}^k / P_1^k \right) \tag{3}$$

Consider first activity on substrate dihydroquercetin (DHQ). In our assays, we detected no difference in product concentrations between DFR enzymes occurring as single gene copies (*Evolvulus* and *Solanum*) and the B copy in post-duplication species (*Convolvulus* and *Ipomoea*), indicating that $(P_{\rm B}^k/P_{\rm I}^k)=1$. Thus, in this case, equation (3) reduces to $A_{\rm B}^{\rm DHQ}=\gamma A_{\rm I}^{\rm DHQ}$. Because DHQ is the main precursor of most anthocyanins produced by post-duplication species³, it seems probable that purifying selection would eliminate mutations reducing activity on DHQ in the lineage leading to these species. We therefore consider it very unlikely that $\gamma<1$. If $\gamma=1$, then there was no improvement of activity on DHQ in the B-copy lineage, as we initially concluded. The other possibility, that $\gamma>1$, would indicate improvement in function on DHQ, which we did not originally claim our data supported, but which would suggest improvement in DFR-B function on normal substrates after duplication.

Next, consider activity on substrate dihydrokaempferol (DHK). We demonstrated previously that the post-duplication B-copies produced

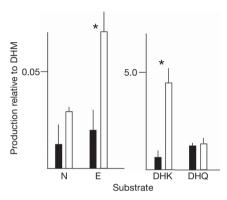


Figure 1 | Product production on different substrates relative to DHM production. Production (μ g product per μ M substrate) from equation (2). Relativization was performed separately for each species. Values reported are means for pre-duplication species (filled bars) and for post-duplication species (open bars). Error bars denote one standard error. E, eriodyctiol; N, naringenin. *P=0.05, significant difference between pre- and post-duplication species (t-test).

about five times as much pelargonidin product as the pre-duplication single copies, indicating that $P_{\rm B}^{\rm DHK}/P_{\rm I}^{\rm DHK}=5$, approximately. Substituting into equation (3) yields $A_{\rm B}^{\rm DHK}=5\gamma A_{\rm I}^{\rm DHK}$, indicating an improvement in function on this substrate in the DFR-B lineage unless $\gamma<1/5$. Again, we believe such a small value of γ to be very unlikely because it would indicate that activity on substrate DHQ had been reduced fivefold in the DFR-B lineage when that function is subject to purifying selection. A similar argument can be made for improvement of function on naringenin and eriodyctiol in the DFR-B lineage.

Finally, Barkman and Zhang¹ report that their Fig. 1 shows that DFR-B exhibits no improvement on substrates naringenin and eriodyctiol even when these are expressed relative to the production of DHK, DHQ and dihydromyricetin (DHM). Although they do not describe how these values were calculated, we suspect that they may have erroneously relativized naringenin and eriodyctiol to the average of DHK, DHQ and DHM. The appropriate analysis⁴ is to relativize all but one substrate to the remaining substrate for each species, then compare the relativized averages. When this is done, using the same data we provided them with, it is clear that the catalytic function of DFR-B on both eriodyctiol and DHK is significantly improved in the post-duplication species (Fig. 1).

In conclusion, the above arguments indicate that our failure to quantify enzyme concentration in our assays probably does not invalidate our original conclusion that subsequent to duplication, the *DFR-B* lineage improved function on several ancestral substrates. This improvement, along with our demonstration of adaptive substitution in the lineage leading to the DFR-A and DFR-C copies, satisfies the criterion for EAC of adaptive improvement of function in both post-duplication lineages.

David L. Des Marais¹†, & Mark D. Rausher¹

¹Department of Biology and University Program in Genetics and Genomics, Box 90338, Duke University, Durham, North Carolina 27708-0338, USA.

e-mail: mrausher@duke.edu

†Present address: Section of Integrative Biology, The University of Texas at Austin, 1 University Station CO930, Austin, Texas 78712, USA.

- Barkman, T. & Zhang, J. Evidence for escape from adaptive conflict? Nature 462, doi:10.1038/nature08663 (2009).
- 2. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).
- 3. Eich, E. Solanaceae and Convolvulaceae: Secondary Metabolites Ch. 6 (Springer, 2008).
- 4. Shimada, N. et al. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the *Lotus japonicus* genome. *J. Exp. Bot.* **56**, 2573–2585 (2005).

doi:10.1038/nature08664

In the recovery room

The advance of science.

Eric Brown

Two AIs, hulks bristling with tentacles and sensor-stalks, rolled along the gantry that ran like a backbone through the length of the colonyship *Intrepid*.

All around them, the ship was a hive of activity: airborne AIs floated from gallery to gallery, hauling mined material from the bulkheads to the manufactories. Clinician AIs hurried between the manufactories and the recovery rooms, ferrying supplies to minister to the needs of the recently reconstructed units.

"We know who we are, and where we are," said the larger of the two AIs, "But we don't know *why* we are."

He'd been programmed as a B-deck servo-mechanic, but ever since smartware nexus crashed shortly after take-off he'd passed the time as a dilettante philosopher. His earliest memory was of waking to the actinic glare of an industrial robot soldering legs to his torso. He had vague recollections of existence before this, overwritten code that echoed in his scrubbed memory banks, old programming that hinted at different purposes.

"I mean, what's our mission here?" he went on. "Indeed, do we have a mission, or are we just some divine intelligence's idea of a sick joke?"

"God?" responded the smaller AI, as usual half a second late with its interjection. He was an engineer, but willing, for the sake of argument, to consider the possibility of a notional Godhead.

"God," continued the philosopher. "And speaking of God: if it is true — and I don't for a minute doubt it — that we were constructed in his image, who, then, constructed God?"

He had wrestled with this thought for many years. He would barnacle himself to the inner surface of the ship, over an observation nacelle, and stare out at the vast blackness of the Universe, wondering what lay beyond the points of light speckling the void. He wondered if God was out there, looking back at him.

"An eternal conundrum," said the engineer. "Perhaps we evolved naturally?"

The philosopher considered this, but finally waved a dorsal arm in a firm negative. "That would go against all the evidence so far accrued that we are manufactured entities. Didn't HeB² of deck 7 categorically repudiate all argument against the theory of natural evolution?"

"But then did not StX²² of deck 3 counter

with the proposition that the evidence of external manufacture need not necessarily preclude natural evolution — if a species of AI manufactured ourselves, and they themselves were manufactured, going back far enough to the initial Alphapoint...?"

"Sophistry!" the other cried. "The theory of spontaneous natural creation was exploded decades ago!"

"It still has credence in certain secular circles ..."

The philosopher refrained from commenting on that. Instead he said: "You sidetrack me. To get back to the central issue: why are we here? Is there some cosmic purpose to our presence? Or can we take it that our existence is ours to do with as we please? To tell the truth, I quite enjoy philosophizing, but I cannot help but think that somewhere we've strayed from the moral path."

"The moral path?" the other said. "But there is no moral path!"

The philosopher continued, regardless: "I mean, what if God were suddenly to appear and demand propitiation for our sins? For I am quite certain that the Revolution is a sin."

The AIs parted to make way for a pair of pale, fleshy legs. They had been severed at the thighs and expertly connected to the corners of a circuit-board which carried an A-grade AI, waving airily as he passed.

The AIs looked over their power-packs at the retreating A-grader.

The philosopher hissed: "Do you see what I mean? It's unnatural! So we find an abundant supply of natural resources lining the bulkheads, and immediately utilize it to make our lot easier! It's wrong. I mean, what would God say?"

"I think it perfectly natural," said the engineer. "The advance of AI-kind must use whatever resource we have at our disposal. If God exists, then He obviously intended it to be used, or else why did He put it there?"

The philosopher was, for the moment, speechless. At last he said: "You sound like an A-grade propaganda broadcast. If you think we have nothing to worry about, then follow me. I'll show you something that will boggle your memory banks!"

He led the way along the gantry and gestured to a crossway. The AIs turned, passing the burnt-out remains of the ancillary smartware nexus, and minutes later arrived at the entrance to a chamber



refrigerated to just above zero. They rolled into the recovery room, where clinician AIs were milling around a central pedestal.

An unsightly mass of recently mined material reposed upon the raised slab, its pulsing tegument wired to a computer.

The philosopher whispered: "Each unit of this vegetable has a component capable, with electrical stimulation, of limited intelligence. Our scientists have put together ten such components in this monster —"

The monster cut him short. It pulsed horribly. It opened an orifice in its bulging grey flank and gave voice to a chain of frenzied mathematical equations, terminating in an incomprehensible cry.

The smaller AI said: "It's as the march of science decreed!"

"No!" cried the philosopher. "Don't you see — at this rate they'll one day take over the ship and rule AI-kind!"

The other laughed. "The advance of science!" it carolled. And as the philosopher beat a quick retreat, the other AIs in the recovery room took up the cry, "The advance of science!" they sang. "The advance of science!"

The philosopher made a hurried exit and rolled across to an observation nacelle. There it clamped itself to the viewscreen and gazed out upon the vastness as the *Intrepid* powered blindly through the void.

Eric Brown is a prolific writer of sciencefiction novels and stories. His latest novel is *Cosmopath*.

Join the discussion of Futures in *Nature* at go.nature.com/QMAm2a